

-1499-

MEL Q KRG+F+ M Y+D NRV++IY +PLAEIV++FFD+LKSST+GYASFDY++ Y+

Sbjct: 429 MELCQKGRGNFIDMQYLDANRVSIIDMPLAEIVYEFFDQLKSSTKGYASFDYELIGYKP 488

Query: 488 SQLVKMDILLNGDKVDALSFIVHKEFAYERGKIIIVEKLKKIIPRQQFEVPIQAAIGQKIV 547

5 S+LVKMDI+LNG+K+DALSFIVH+++AYERGK+IVEKLK++IPRQQFEVP+QAAIGQKIV

Sbjct: 489 SKLVKMDIMLNGEKIDALSFIVHRDYAYERGKVIVEKLKELIPRQQFEVVPQAAIGQKIV 548

Query: 548 ARSDIKALRKNNVLAKCYGGDVSRKRKLEKQKAGKRRMKAIGSVEVPQEAFLSVLSMDDD 607

ARS IKA+RKNVLAKCYGGD+SRKRKLEKQK GK+RMK +GSVEVPQEAFLSVLSMDD

10 Sbjct: 549 ARSTIKAMRKNNVLAKCYGGDISRKRKLEKQKEGKRRMKQVGSVEVPQEAFLSVLSMDD 608

Query: 608 TKK 610

KK

Sbjct: 609 PKK 611

15

An alignment of the GAS and GBS proteins is shown below.

Identities = 587/610 (96%), Positives = 601/610 (98%)

Query: 1 MNIEDLKKRQEKIRNFSIIAHIDHGKSTLADRILEKTETVSSREMQAQLLDSMDLERERG 60

20 MN +DLKKRQEKIRNFSIIAHIDHGKSTLADRILEKTETVSSREMQAQLLDSMDLERERG

Sbjct: 1 MNSQDLKKRQEKIRNFSIIAHIDHGKSTLADRILEKTETVSSREMQAQLLDSMDLERERG 60

Query: 61 ITIKLNAIELNLTAKDGETYIFHLIDTPGHVDFTYEVSRLAACEGAILVVDAAQGIEAQ 120

ITIKLNAIELNLTAKDGETYIFHLIDTPGHVDFTYEVSRLAACEGAILVVDAAQGIEAQ

25 Sbjct: 61 ITIKLNAIELNLTAKDGETYIFHLIDTPGHVDFTYEVSRLAACEGAILVVDAAQGIEAQ 120

Query: 121 TLANVYLALDNDLEILPVINKIDLPAADPERVRAEVEDVIGLDASEAVLASAKAGIGIEE 180

TLANVYLALDNDLEILPVINKIDLPAADPERV EVEDVIGLDASEAVLASAKAGIGIEE

30 Sbjct: 121 TLANVYLALDNDLEILPVINKIDLPAADPERVRHEVEDVIGLDASEAVLASAKAGIGIEE 180

Query: 181 ILEQIVEKVPAPTGEVDAPLQALIFDSVYDAYRGVILQVRIVNGMVKPGDKIQMMSNGKT 240

ILEQIVEKVPAPTG+VDAPLQALIFDSVYDAYRGVILQVRIVNG+VKPGDKIQMMSNGKT

Sbjct: 181 ILEQIVEKVPAPTGDVDAPLQALIFDSVYDAYRGVILQVRIVNGIVKPGDKIQMMSNGKT 240

35 Query: 241 FDTVTEVGIFTPKAVGRDFLATGDVGYYAASIKTVADTRVGDTITLANNPAIEPLHGYKQM 300

FDTVTEVGIFTPKAVGRDFLATGDVGYYAASIKTVADTRVGDT+TLANNPA E LHYGYKQM

Sbjct: 241 FDTVTEVGIFTPKAVGRDFLATGDVGYYAASIKTVADTRVGDTVTLANNPAKEALHGYKQM 300

Query: 301 NPMVFAGLYPIESNKYNDLREALEKLQNDASLQFEPETSQALGFGFRGFLGLLHMDVI 360

40 NPMVFAG+YPIESNKYNDLREALEKLQNDASLQFEPETSQALGFGFRGFLGLLHMDVI

Sbjct: 301 NPMVFAGIYPIESNKYNDLREALEKLQNDASLQFEPETSQALGFGFRGFLGLLHMDVI 360

Query: 361 QERLEREFNIDLIMTAPSVVYHVNTDGEMLVSNPSEFPDPTRVDSIEEPYVKAQIMVP 420

QERLEREFNIDLIMTAPSVVYHV+TTD +M+EVSNPSEFPDPTRV IEPPYVKAQIMVP

45 Sbjct: 361 QERLEREFNIDLIMTAPSVVYHVHTDDEDMIEVSNPSEFPDPTRVAFIEPPYVKAQIMVP 420

Query: 421 QEFVGAVMELAQKRKRGDFVTMDYIDNVRNVIIYQIPLAEIVDFDFDKLKSSTRGYASFDY 480

QEFVGAVMEL+QKRKRGDFVTMDYIDNVRNVIIYQIPLAEIVDFDFDKLKSSTRGYASFDY

50 Sbjct: 421 QEFVGAVMELSQRKRGDFVTMDYIDNVRNVIIYQIPLAEIVDFDFDKLKSSTRGYASFDY 480

Query: 481 EISEYRRSQLXKMDILLNGDKVDALSFIVHKEFAYERGKLIIVDKLKKIIPRQQFEVPIQA 540

++SEYRRSQL KMDILLNGDKVDALSFIVHKEFAYERGK+IV+KLKKIIPRQQFEVPIQA

Sbjct: 481 DMSEYRRSQLVKMDILLNGDKVDALSFIVHKEFAYERGKIIIVEKLKKIIPRQQFEVPIQA 540

55 Query: 541 AIGQKIVARSDIKALRKNNVLAKCYGGDVSRKRKLEKQKAGKRRMKAIGSVEVPQEAFLS 600

AIGQKIVARSDIKALRKNNVLAKCYGGDVSRKRKLEKQKAGKRRMKAIGSVEVPQEAFLS

Sbjct: 541 AIGQKIVARSDIKALRKNNVLAKCYGGDVSRKRKLEKQKAGKRRMKAIGSVEVPQEAFLS 600

Query: 601 VLSMDDDDKK 610

VLSMDDD KK

60 Sbjct: 601 VLSMDDDTKK 610

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1500-

Example 1370

A DNA sequence (GBSx1455) was identified in *S.agalactiae* <SEQ ID 4189> which encodes the amino acid sequence <SEQ ID 4190>. This protein is predicted to be awd gene product (ndk). Analysis of this protein sequence reveals the following:

```

5   Possible site: 42
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2097 (Affirmative) < succ>
10   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAF57188 GB:AE003779 awd gene product [Drosophila melanogaster]
    Identities = 73/136 (53%), Positives = 100/136 (72%), Gaps = 5/136 (3%)

    Query: 2   EQTFFMIKPDGVKRGFGEVISRIERRGFSIDRLVRYADADILKRHYAELTDRPFPPTL 61
              E+TF M+KPDGV+RG +G++I R E++GF + L+ +A ++L++HYA+L+ RPFFP L
    Sbjct: 25  ERTFIMVKPDGVQGRGLVGKTIERFEQKGFKLVALKFTWASKELLEKHYADLSARPFPFGL 84

20   Query: 62  VDYMTSGPVIIGVISGEEVISTWRIMMGSTNPKDALEPGTIRGDFQAQAPSPNQATCNIVHG 121
              V+YM SGPV+ V G V+ T R M+G+TNP D+LEPGTIRGDF Q NI+HG
    Sbjct: 85  VNYMNSGPVVPVMVWGLNVVKTRQMLGATNPADSLPGTIRGDFC-----IQVGRNIHG 139

25   Query: 122 SDSPESATREIAIWFN 137
              SD+ ESA +EIA+WFN
    Sbjct: 140 SDAVESAEKEIALWFN 155

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4191> which encodes the amino acid sequence <SEQ ID 4192>. Analysis of this protein sequence reveals the following:

```

30   Possible site: 22
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
35   bacterial cytoplasm --- Certainty=0.2913 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

40   Identities = 30/48 (62%), Positives = 35/48 (72%)

    Query: 87  MMGSTNPKDALPGTIRGDFQAQAPSPNQATCNIVHGS DSPESATREIAI 134
              MM TNPKDAL GTIR +FAQAP + N+VHGS S +SA REIA+
45   Sbjct: 1   MMRVTNPKDALCGTIRENFAQAPGDDGGIFNMVHGSHSRDSARREIAL 48

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1371

A DNA sequence (GBSx1456) was identified in *S.agalactiae* <SEQ ID 4193> which encodes the amino acid sequence <SEQ ID 4194>. Analysis of this protein sequence reveals the following:

```

50   Possible site: 15
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
55   bacterial cytoplasm --- Certainty=0.2734 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

```

-1501-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4195> which encodes the amino acid sequence <SEQ ID 4196>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1985(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 22/34 (64%), Positives = 26/34 (75%)

Query: 28 SFGTIRNSTALKQLTLDLNLNLSFGTIRNSTALK 61
SFGTI+NS ALKQ + +N SFGTI+NS ALK

Sbjct: 7 SFGTIQNSIALKQKAQEENQRSFGTIQNSIALK 40
Identities = 22/34 (64%), Positives = 26/34 (75%)

Query: 6 SFGTIRNSTALKLYAKQSPAFRSFGTIRNSTALK 39
SFGTI+NS ALK A++ RSFGTI+NS ALK

Sbjct: 7 SFGTIQNSIALKQKAQEENQRSFGTIQNSIALK 40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1372

A DNA sequence (GBSx1457) was identified in *S.agalactiae* <SEQ ID 4197> which encodes the amino acid sequence <SEQ ID 4198>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1407(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4199> which encodes the amino acid sequence <SEQ ID 4200>. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2055(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 154/221 (69%), Positives = 187/221 (83%)

Query: 1 MIKINFPILDEPLVLSNATILTIEDVSVYSSLVKHFYQYDVDEHLKLFDDKQKSLKATEL 60
++ +NF +LDEP+ L TIL +EDV V+S +V++ YQY+ D LK FD K K++K +E+

-1502-

Sbjct: 8 LMNLFNSLLDEPIPLRGGTILVLEDDVCVFSKIVQCYQYEDSELKFFDHKMKTIKSEI 67

Query: 61 MLVTDILGYDVNSAPILKLIHGDLNQFNEKPEVKSMVEKLAATITELIAFECLLENELDL 120
MLVTDILG+DVNS+ ILKLIH DLE+QFNEKPEVKSM++KL ATITELI FECLLENELDL

5 Sbjct: 68 MLVTDILGFDVNSSTILKLIHADLESQFNEKPEVKSMIDKLVATITELIVFECLLENELDL 127

Query: 121 EYDEIKILELIKALGVKIETQSDTIFEKCFEIIQVYHYLTKKNLLVFNVSAGAYLTKDEVI 180
EYDEI ILELIK+LGVK+ETQSDTIFEK EI+Q++ YLTKK LL+FNVSAG+LTKDEV

10 Sbjct: 128 EYDEITILELIKSLGVKQVETQSDTIFEKCLEILQIFKYLTKKLLIFVNSGAFITKDEVA 187

Query: 181 KLCEYINLMQKSVLFLEPRRLYDLPOYVIDKDYFLIGENMV 221
L EYI+L +VLFLEPR LYD PQY++D+DYFLI +NMV

Sbjct: 188 SLQEYISLTNLTVLFLPRELYDFFQYILDEDYFLITKNMV 228

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1373

A DNA sequence (GBSx1458) was identified in *S.galactiae* <SEQ ID 4201> which encodes the amino acid sequence <SEQ ID 4202>. Analysis of this protein sequence reveals the following:

20 Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.0842(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9783> which encodes amino acid sequence <SEQ ID 9784> was also identified.

- 30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB83918 GB:AL162753 hypothetical protein NMA0629 [Neisseria
meningitidis Z2491]
Identities = 45/104 (43%), Positives = 65/104 (62%), Gaps = 2/104 (1%)

35 Query: 4 RYMRMILMFDMPETAEERKAYRKFRKFLLESEGFMHQFSVYSKLLNNTANNAMIGRLK 63
++MR+I+ FD+P TA +RKA +FR+FLL +G+ M Q SVYS+++ + RL

Sbjct: 5 KFMRIIVFFDLPVITAARKKAANQFRQFLKDGQMLQLSVYSRIVKGRDSLQKHNRLC 64

40 Query: 64 VNNPKKGNITLLTVTEKQFARMVYLHGERNT--SVANSDSRLVF 105
N P++G+I L +TEKQ+A M L GE T NSD L+P

Sbjct: 65 ANLPQEGSIRCLEITEKQYAAMKLLLGELKTQEKVNSDQLLLF 108

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4203> which encodes the amino acid sequence <SEQ ID 4204>. Analysis of this protein sequence reveals the following:

45 Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.0822(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 97/112 (86%), Positives = 107/112 (94%)

55 Query: 1 MSYRYNRMILMFDMPETAEERKAYRKFRKFLLESEGFMHQFSVYSKLLNNTANNAMIG 60
MSYRYNRMILMFDMPET+TAEERKAYRKFRKFLLESEGFMHQFS+YSKLLNNTANNAMIG

-1503-

Sbjct: 1 MSYRYMRMILMFDMPTDTAERKAYRKFRKFLLESEGFMHQFSIYKLLNNTANNAMIG 60

Query: 61 RLKVNPNPKGNITLLTVTEKQFARMVYLHGERNTSVANSDSLVLGDSYDQ 112
 RL+ +NP KGNITLLTVTEKQFARM+YHGERN +ANS DLVLG+++D+

5 Sbjct: 61 RLREHNPKNKGNITLLTVTEKQFARMYIYHGERNNCIANSRDLVLGEAFDE 112

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1374

10 A DNA sequence (GBSx1459) was identified in *S.galactiae* <SEQ ID 4205> which encodes the amino acid sequence <SEQ ID 4206>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3185(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB83919 GB:AL162753 hypothetical protein NMA0630 [Neisseria
 meningitidis Z2491]
 Identities = 71/224 (31%), Positives = 122/224 (53%)

25 Query: 4 WRTVVVNTHSKLSYKNNHLIFKDSYQTEMIHLSEIDILIMETTDIVLSTMLIKRLVDENI 63
 WR++++ KLS + L+ + + + + L +I ++I+E + + + L+ L +
 Sbjct: 3 WRSLLIQNGGKLSLQRRQLLIQONGESHVPLEDIAVIIENRETLITAPLLSALAEHGA 62

30 Query: 64 LVIFCDDKRLPTAMLPYYARHDSSLQLSRQMSWIEDVKADVWTSIIAQKILNQSFYLGE 123
 ++ CD++ LP +PY H L Q++ E +K +W I+ QKILNQ+F E
 Sbjct: 63 TLLTCDEQFLPCGQWLFPYAQYHRQLKILKQLNISEPLKKQLWQHIVRQKILNQAFVADE 122

35 Query: 124 CSFFEKSQSIMNLYHDLEPDPSPNREGHAARIYFNTLFGNDFSREQDNPINAGLDYGYSL 183
 ++ + L ++ D NRE AA +YF LFG F+R +N +NA L+Y Y++
 Sbjct: 123 TGNDLAAKRLRTLASEVRSRGDTGNREAQAALYFQALFGEKFTRNDNNAVNAALNYTYAV 182

Query: 184 LLSMFAREVVKCGCMTQFGLKHANQFNQNLASDIMEPPFRPIVD 227
 L + AR + G + GL H ++ N FNLA D +EP RP+ D
 Sbjct: 183 LRAAVARALTLYGWLPLGLFHRSELPNPNLADDFIEPLRLPAD 226

40

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4207> which encodes the amino acid sequence <SEQ ID 4208>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

45

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3185(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

An alignment of the GAS and GBS proteins is shown below.

Identities = 239/289 (82%), Positives = 271/289 (93%)

55 Query: 1 MAGWRTVVVNTHSKLSYKNNHLIFKDSYQTEMIHLSEIDILIMETTDIVLSTMLIKRLVD 60
 MAGWRTVVVNTHSKLSYKNNHLIFKD+Y+TE+IHLSEIDIL++ETTDIVLSTML+KRLVD
 Sbjct: 1 MAGWRTVVVNTHSKLSYKNNHLIFKDAYKTEL IHLSEIDILLETTDIVLSTMLVKRLVD 60

Query: 61 ENILVIFCDDKRLPTAMLPYYARHDSSLQLSRQMSWIEDVKADVWTSIIAQKILNQSFY 120
 EN+LVIFCDDKRLPTAMLP+Y RHDSSLQL +QMSW E VK+ VWT+IIAQKILNQS Y

-1504-

5
10
15

Sbjct: 61 ENVLVIFCDKRLPTAMLMPFYGRHDSLSLQLGKQMSWSETVKSQVWTTIIAQKILNQSCY 120

Query: 121 LGECSFFEKSQSIMNLYHDLEFPDPSNREGHAARIYFNTLFGNDFSRQDNPINAGLDYG 180
LG CS+FEKSQSIM+LYH LE FDPNREGHAARIYFNTLFGNDFSR+ ++PINAGLDYG

Sbjct: 121 LGACSYFEKSQSIMDLVHGLENFDPNREGHAARIYFNTLFGNDFSRDLHPINAGLDYG 180

Query: 181 YSLLLSMFAREVVKCGCMTQFGLKHANQFNQFNLASDIMEPPRPIDRIIYENRQSDFKV 240
Y+LLLSMFAREVV GCMTQFGLKHANQFNQFN ASDIMEPPRP+VD+I+YENR F K

Sbjct: 181 YTLLLSMFAREVVVSGCMTQFGLKHANQFNQFNASDIMEPPRPLVDKIVYENRNQFFPK 240

Query: 241 MKRELFSMFSETYSYNGKEMYLNI+SDYTKKV+K+LN++G G+PEFRI 289
+KREL++FS+T+SYNGKEMYL+NI+SDYTKKV+K+LN++G G+PEFRI

Sbjct: 241 IKRELFTLFSDFSYNGKEMYLNI+SDYTKKV+K+LN++G G+PEFRI 289

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1375

A DNA sequence (GBSx1460) was identified in *S.agalactiae* <SEQ ID 4209> which encodes the amino acid sequence <SEQ ID 4210>. Analysis of this protein sequence reveals the following:

20 Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.1109(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:CAB73943 GB:AL139078 hyopthetical protein Cj1523c [Campylobacter jejuni]
Identities = 165/746 (22%), Positives = 291/746 (38%), Gaps = 115/746 (15%)

35 Query: 318 LSASMIQRYDEHREDLKQLKQFVKASLPEKYQEI--FADSSKDGAGYIEGKTNQEAFFK 375
L+ S +R + L LK + Y++ F+S Y G + E ++

Sbjct: 50 LARSARKRLARRKARLNHLKHLIANEFKLNIEDYQSFDES LAKAYKGLISP--YELRFR 107

40 Query: 376 YLSKLLTKQEDSENFL--KIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRS----- 428
L++LL+KQ+ + L K + D ++ + G+I + E K + QS

Sbjct: 108 ALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEK-LANYQSVGEYL 166

45 Query: 429 --EYYPFLKENQDRIEKILTFRIPIY-----IGPLAREKSDFAW-MTRKTDDSI 474
EY+ KEN + + Y + + +++ +F + ++K ++ +

Sbjct: 167 YKEYFQKPKENSKEFTNVRNKKESYERCIASFLKDELKLIFKKQREFGFSFKKFEEV 226

50 Query: 475 RPWNFEDLVDEKSAEAFIHRMTNNDFFLPEEKVLPKHS LIYEKFTVYNELTKV--RYKN 532
F +++ + FH + N F+ +EK PK+S + F + + KN

Sbjct: 227 LSVAFY-----KRALKDFSHLVGNCSSFFT-DEKRAPKNSPLAFMFVALTRIINLNNLKN 280

55 Query: 533 EQGETYFFDSNIKQEIFDGVFKEHRKVS--KKLLDFLAKEYEFRIVDVIGLDKENKAF 590
+G Y D + + V K K KLL L+ +YE E +

Sbjct: 281 TEGILYTKDD--LNALLNEVLKNGTILTYKQTKLLG-LSDDYE-----FKGEGKTY 328

60 Query: 591 NASLGTYHDLKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLNENYKDLFTESQL 650
Y + K L + L D L +I + +TL +D +KK L Y ++Q+

Sbjct: 329 FIEFKKYKEFIKALGEHNSQDD---LNEIAKDITLIKDEIKLKKALAKYD--LNQNQI 382

Query: 651 KKLYRRHYTGWGRLSAKLINGIRDK--ESQKTILDYLDGSRNRFMQLINDDGLSFKS 708
L + + +S K + + E +K D+ + N IN+D F

Sbjct: 383 DLSKLEFKDHLNISFKALKLVTPLMLEGKK-----YDEACNELNLKVAINEDKKDFLP 436

Query: 709 IISKAQAGSHSDNLKEVVGELAGSPAIAKKGILQSLKIVDELVKVMGYEPEQIVVEMAREN 768
++ N P + + I + K+++ L+K G + +I +E+ARE

-1505-

Sbjct: 437 AFNETYYKDEVIN-----PVVLRAIKEYRKVLNALLKKYG-KVHKINIELAREV 484

Query: 769 QTTNQGR----RNSRQRYKLDDG---VKNLASDLNG-NILKEYPTDNQALQNERLFLYY 820
 + R + + YK D + L +N NILK L L+

5 Sbjct: 485 GKNHSQRAKIEKEQENENYKAKKDAELECEKLGKINSKNILK-----LRLFK 531

Query: 821 LQNGRDMYTGEALDIDNLSQ---YDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPS 877
 Q Y+GE + I +L +IDHI P + DDS N+VLV + +N+ K + P

10 Sbjct: 532 EQKEFCAYSGBEKIKISDLQDEKMLEIDHIYPYSRSDDSYMNKVLVFTKQNEKLNQTP- 590

Query: 878 LEIVKDCVKVFWKKL--LDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRLVETRQITKH 935
 E + W+K+ L L +++++ L K ++ F R L +TR I +

Sbjct: 591 FEAFGNSAKWQKIEVLAKNLPTKKQKRILDK---NYKDKEQKNFKDRNLNDTRYIARL 646

15 Query: 936 VARI-----LDERFNNELDSKGRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNY 986
 V L + N +L+ ++ KV + L S R +GF N+

Sbjct: 647 VLNYTKDYLDLFLPLSDDENTKLNDT-QKGSKVHVEAKSGLTSALRHWTGFSAKDRNNHL 705

Query: 987 HHAHDAYLNAVVAKAILTKYPQLEPE 1012
 HHA DA + A +I+ + + E

20 Sbjct: 706 HHAIDAVIIAYANNSIVKAFSDFKKE 731

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4211> which encodes the amino acid sequence <SEQ ID 4212>. Analysis of this protein sequence reveals the following:

25 Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.0973(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 881/1380 (63%), Positives = 1088/1380 (78%), Gaps = 22/1380 (1%)

35 Query: 1 MNKPYSIGLDIGTNSVGSIIITDDYKVPKAKMRVLGNTDKEYIKKNLIGALLFDGGNTAA 60
 M+K YSIGLDIGTNSVGW++ITD+YKVP+KK +VLGNTD+ IKKNLIGALLFD G TA
 Sbjct: 1 MDKYSIGLDIGTNSVGWAVITDEYKVPKSKKFKVLGNTDRHSIKKNLIGALLFDGSGTAE 60

40 Query: 61 DRRLKRTARRRYTRRRNRILYLQEIFAEMSKVDDSFHRLSDSFLVEEDKRGSKYPIFA 120
 RLKRTARRRYTRR+NRI YLQEIF+ EM+KVDDSFHRLS+SLVEEDK+ ++PIF
 Sbjct: 61 ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSESLVEEDKKHERHPIFG 120

45 Query: 121 TLQEEKDYHEKFSTIYHLRKLADKKEKADLRLLIYIALAHIIKFRGHFLIEDDSFDVRNT 180
 + +E YHEK+ TIYHLRK+L D +KADLRLLY+ALAH+IKFRGHFLIE D + N+
 Sbjct: 121 NIVDEVAYHEKYPTIYHLRKLVDSTDKADLRLLIYIALAHMIKFRGHFLIEGD-LNPDNS 179

Query: 181 DISKQYQDFLEIFNTTFENNLLSQNVDEVAILTDKISKSAKKDRILAQYPNQKSTGIFA 240
 D+ K + ++ +N FE N + + VD +AIL+ ++SKS + + ++AQ P +K G+F

50 Sbjct: 180 DVDKLFQILVQTYNQLFEENPINASGVDAKILSARLSKSRLENLIAQLPGEKKNGLFG 239

Query: 241 EFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLNLLGQIGDEFADLFSAAKKLYDS 300
 + L +G +FK F+L + LQ +KD+YD+DL+NLL QIGD++ADLF AAK L D+

55 Sbjct: 240 NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDA 299

Query: 301 VLLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDG 360
 +LLS IL V TKAPLSASMI+RYDEH +DL LK V+ LPEKY+EIF D SK+GY

Sbjct: 300 ILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY 359

60 Query: 361 AGYIEGKTNQEAIFYKLSKLLTKQEDSENFLKIKNEDFLRKQRTFDNGSIPHQVHLTEL 420
 AGYI+G +QE FYK++ +L K + +E L K+ ED LRKQRTFDNGSIPHQ+HL EL

Sbjct: 360 AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL 419

65 Query: 421 KAIIRROSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTRKTDSDIRPWNFE 480
 AI+RRQ ++YPFLK+N+++IEKILTFRIPYY+GPLAR S FAWMTRK+++I PWNFE

-1506-

Sbjct: 420 HAILRRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAMWTRKSEETITPWNFE 479
 Query: 481 DLVDKEKSAAEAFIHRMTNDFYLPPEEKVLPKHSLLIYEKFTVYNELTKVRYKNE-QGETYF 539
 ++VDK SA++FI RMTN D LP EKVLPKHSLL+YE FTVYNELTKV+Y E + F
 5 Sbjct: 480 EVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAF 539
 Query: 540 FDSNIKQEIFDGVFKEHRKVSKKLLDFLAKEYEDEFRIVDVIGLDKENKAFNASLGTYHD 599
 K+ I D +FK +RKV+ K+L + K+ E F V++ G++ FNASLGTYHD
 10 Sbjct: 540 LSGEQKKAIVDLLFKTRNKVTVKQLKEDYFKKIECFDSVEISGVEDR---FNASLGTYHD 596
 Query: 600 LEKIL-DKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLNENYKDLFTESQLKKLYRRHY 658
 L KI+ DKDFLDN +NE ILEDIV TLTFLFEDREMI++RL+ Y LF + +K+L RR Y
 Sbjct: 597 LLKIIKDKDFLDNEENEDILEDIVLTTLTLFEDREMIERLKYAHLFDDKVMKQLKRRY 656
 15 Query: 659 TGWGRLSAKLINGIRDKESQKTILDYLIIDGRSNRNFQMLINDDGLSFKSIISKAQAGSH 718
 TGWGRLS KLINGIRDK+S KTILD+L DG +NRNFQMLI+DD L+FK I KAQ
 Sbjct: 657 TGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAVSGQ 716
 Query: 719 SDNLKEVVGELAGSPAIIKKGILQSLKIVDELVKVMG-YEPEQIVVEMARENQTTNQGRN 777
 D+L E + LAGSPAIIKKGILQ++K+VDELVKVMG ++PE IV+EMARENQTT +G++N
 20 Sbjct: 717 GDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVEMARENQTTQKGQKN 776
 Query: 778 SRQRYKLDDGVKNLASDLNGLNKEYPTDNQALQNERLFLYYLQNGRDMYTGEALDIDN 837
 SR+R K +++G+K L S ILKE+P +N LQNE+L+LYYLQNGRDMY + LDI+
 25 Sbjct: 777 SRERMKRIEIGIKELGS---QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR 832
 Query: 838 LSQYDIDHIIPQAFIKDDSIDNRVLSSAKNRGKSDDVPSLEIVKDCVFWKKLLDAKLM 897
 LS YD+DHI+PQ+F+KDDSIDN+VL S KNRGKSD+VPS E+VK K +W++LL+AKL+
 30 Sbjct: 833 LSDYDVHIVPQSFLKDDSIDNKVLTSDKNRGKSDNVPSEEVKKMKNYWRQLLNAKLI 892
 Query: 898 SQRYKYNLTKAERGGLTSDDKARFIQRLVETRQITKHVARILDERFNNELDSKGRIRK 957
 +QRK+DNLTKAERGGL+ DKA FI+RQLVETRQITKHVA+ILD R N + D + IR+
 Sbjct: 893 TQRKFDNLTKAERGGLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIRE 952
 35 Query: 958 VKIVITLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVAKAILTKYPQLEPEFVYGD 1017
 VK++TLKS LVS+FRK+F FYK+RE+NNYHHAHDAYLNAV A++ KYP+LE EFVYGD
 Sbjct: 953 VKVITLKSGLVSDFRKDFQFYKREINNYHHAHDAYLNAVGTALIKKYPKLESEFVYGD 1012
 Query: 1018 YPKYN-----SYKTRKSATEKLFYFSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEI 1070
 Y Y+ S + AT K FFYSNIMNFFKT++TLA+G + + IE N +TGEI
 40 Sbjct: 1013 YKVYDVRKMIKSEGEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRLIETNGETGEI 1072
 Query: 1071 VWDKKKHFAFVRKVLSTPQNNIVKTEIQTGGFSKESILAHGNSDKLI PRKTKDIYLDPK 1130
 VWDK + FATVRKVLST PQ NIVKTE+QTGGFSKESIL NSDKLI RK KD DPK
 45 Sbjct: 1073 VWDKGRDFATVRKVLSTPQNNIVKTEVQTGGFSKESILPKRNSDKLIARK-KD--WDPK 1129
 Query: 1131 KYGGFDSPIVAYSVLVADIKKGAQKLTVTTELLGITIMERSRFEKNPSAFLESKGYLN 1190
 KYGGFDSP VAYSVLVVA ++KGK++KLK+V ELLGITIMERS FEKNP FLE+KG Y
 50 Sbjct: 1130 KYGGFDSPTVAYSVLVAKVEKGKSKLLKSVKELLGITIMERSSEKNPIDFLEAKGYKE 1189
 Query: 1191 IRADKLIILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLASRYNESKGP 1250
 ++ D +I LPKYSLFELENGR+R+LASAGELQKGNELALP++++ FLYLAS Y + KG P
 Sbjct: 1190 VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLAHYEKLKGP 1249
 55 Query: 1251 EEIEKKQEFVNQHSYFDDILQLINDFSKRVLADANLEKINKLYQDNKENISVDELANN 1310
 E+ E+KQ FV QH Y D+I++ I++FSKRVLADANL+K+ Y +++ + E A N
 Sbjct: 1250 EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLKDVL SAYNKHDRK-PIREQAEN 1308
 Query: 1311 IINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSIITGLYETRIDLGLGED 1370
 II+LFT T+LGAPAAFK+FD +DRKRYTSTKEVL++TLIHQSIITGLYETRIDL +LG D
 60 Sbjct: 1309 IIHFLTTLNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLGGD 1368

SEQ ID 4210 (GBS317) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell
 extract is shown in Figure 27 (lane 2; MW 179.3kDa) and in Figure 159 (lane 5 & 6; MW 180kDa). It was
 65 also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure
 27 (lane 3; MW 154.3kDa) and in Figure 159 (lane 9 & 10; MW 154kDa).

-1507-

GBS317-GST was purified as shown in Figure 224, lane 9-10. GBS317-His was purified as shown in Figure 222, lane 9.

GBS317N was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 149 (lane 2-4; MW 116kDa).

5 GBS317C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 166 (lane 6-8; MW 92kDa).

GBS317dN was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 187 (lane 7; MW 116kDa). Purified GBS317dN-GST is shown in Figure 245, lane 8.

10 GBS317C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 188 (lane 13; MW 92kDa). Purified GBS317dC-GST is shown in Figure 245, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1376

15 A DNA sequence (GBSx1461) was identified in *S.agalactiae* <SEQ ID 4213> which encodes the amino acid sequence <SEQ ID 4214>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -11.94    Transmembrane    132 - 148 ( 123 - 156)
    INTEGRAL    Likelihood = -11.09    Transmembrane    190 - 206 ( 183 - 209)
20    INTEGRAL    Likelihood = -4.94    Transmembrane    95 - 111 ( 94 - 115)

----- Final Results -----
        bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related sequence was also identified in GAS <SEQ ID 9133> which encodes the amino acid sequence <SEQ ID 9134>. Analysis of this protein sequence reveals the following:

```
30    Possible site: 22

>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -7.32    Transmembrane    126 - 142
    INTEGRAL    Likelihood = -6.90    Transmembrane    178 - 194
35

----- Final Results -----
        bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 94/204 (46%), Positives = 139/204 (68%)

Query: 5    LMKDKLLVVLTVIWIISLATLATIYIAWLIYPIETIQFLKLEKVVYLKAETIYYNFNKLMI 64
45    +M +   ++ +W+W+++LA L TIY WL YP+E+ LKLE+VV++ + I +N+N L+
Sbjct: 4    VMVENTKLLCSWVWLLALAILITIYSTWLWYPLEVDHLKLEQVVFMSKDAILHNYNGLLN 63

Query: 65    YLTHPFISDLNMPSEFPSSDGLKHFADVKYLFTHAGLGFVILTFPVIYFLRRGWKQKSIF 124
        YLT+PF++ L   +F SS DGLKHFADVK+LF L   +F+ L +P +   + K K +
50    Sbjct: 64    YLTNPFVTRLEFANFHSSADGLKHFADVKWLFHLTQVVFLGLLYPTLKTFTQRLKTKRFW 123
```

-1508-

Query: 125 L YEGFFKIAIMLP I FIVVCAFL LGFDQFFTLFHEVLFP GDSTWQFNPLTDPVIWILPETF 184
 L + +A + P+ I + A +GF+ FFTLFH+VLF GDS+W F+PL D VIWILPE F
 Sbjet: 124 LLQKPLILALFPLMIGLMASFIGFEHFFTLFHQVLFVGDSSWLFDPKDSVIWILPEVF 183

Query: 185 FLHCFIIFLLIYETITITILLIIGR 208
 FLHCF+ F+++YE I L+ + R
 Sbjet: 184 FLHCFLFMIVYEIILWSLVGLAR 207

10 SEQ ID 4214 (GBS167) was expressed in and purified from *E.coli*. The purified protein is shown in lanes 5 & 6 of Figure 223.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1377

15 A DNA sequence (GBSx1462) was identified in *S.galactiae* <SEQ ID 4217> which encodes the amino acid sequence <SEQ ID 4218>. This protein is predicted to be p-nitrophenyl phosphatase (pho2). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3925(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15219 GB:Z99120 similar to N-acetyl-glucosamine catabolism
 [Bacillus subtilis]
 Identities = 121/249 (48%), Positives = 172/249 (68%)

30 Query: 3 YKGYLIDLDGTIYKGKSRIPAGERFIERLQKGI PYMLVTNNITRTPESVQEMLRGFNVE 62
 YKGYLIDLDGT+Y G +I F+ L+++G+PY+ VTNN++RTP+ V + L F++
 Sbjet: 4 YKGYLIDLDGTMNGTEKIEEACEFVRTLKDRGV PYLEVTNNSSRTPKQVADKLVSFDP 63

35 Query: 63 TPLETTYTATMATVDYMDMNRGKTAYVIGEEGLKKAIDAGYVEDTKNPAYVVVGLDWN 122
 E ++T +MAT ++ + + YVIGEEG+++AI + G +N +VVVG+D +
 Sbjet: 64 ATEEQVFTTSMATAQHIAQQKSDASVYVIGEEGRQAIEENGLTFGENADFVVVGIDRS 123

40 Query: 123 VTYDKLATATLAIQNGALFIGTNPDLNIPTERGLLPAGSLNALLEAATRIKPVFIGKPN 182
 +TY+K A LAI+NGA FI TN D+ IPTERGLLPG GSL ++L +T ++PVFIGKP
 Sbjet: 124 ITYEKFAVGCLAIRNGARFISINGDIAIPTERGLLPNGSLTSLVTVSTGVQPVFIGKPE 183

45 Query: 183 AIIMNKALEILNIPRNOAVMVGDNLYLTDIMAGINNDITLLVTTGFTTVEEVPDLPIQPS 242
 +IIM +A+ +L ++ +MVGDN Y TDIMAGIN +DTLLV TG T E + D +P+
 Sbjet: 184 SIIMEQAMRVLGTDVSEILMVGDN YATDIMAGINAGMDTLLVHTGVTKREHMTDDMEKPT 243

Query: 243 YVLASLDEW 251
 + + SL EW
 Sbjet: 244 HAIDSLTEW 252

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4219> which encodes the amino acid sequence <SEQ ID 4220>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

55 INTEGRAL Likelihood = -0.53 Transmembrane 128 - 144 (128 - 144)

----- Final Results -----
 bacterial membrane --- Certainty=0.1213(Affirmative) < succ>

-1509-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:CAB15219 GB:Z99120 similar to N-acetyl-glucosamine catabolism
 [Bacillus subtilis]
 Identities = 121/250 (48%), Positives = 166/250 (66%), Gaps = 1/250 (0%)

10 Query: 3 YKGYLIDLDGTIYQGKNRIPAGERFIKRLQERGIPYLLVTNNTRTPPEMVQSMANQPHV 62
 YKGYLIDLDGT+Y G +I F++ L++RG+PYL VTNN++RTP+ V L + F +
 Sbjct: 4 YKGYLIDLDGTMVNGTEKIEEACEFVRTLKDRGVPYLFVTNNSSRTPKQVADKLVN-FDI 62

15 Query: 63 ETSIETIYTATMATVDYMDNMNRGKTAYVIGETGLKSAIAAAGYVEELENPAYVVVGLDS 122
 + E ++T +MAT ++ + + YVIGE G++ AI G EN +VVVG+D
 Sbjct: 63 PATEEQVFTTSMATAQHIAQQKDKASVYVIGEEGIRQAIEENGLTFGGENADFFVVVGIDR 122

20 Query: 123 QVTYEMLAIAATLAIQKGFALFIGTNPDLNIPTERGLMPGAGALNALLEAATRVKPVFIGKP 182
 +TYE A+ LAI+ GA FI TN D+ IPTERGL+PG G+L ++L +T V+PVFIGKP
 Sbjct: 123 SITYEKFAVGCLAIRNGARFISTNGDIAIPTERGLLPNGSLTSVLTIVSTGVQPVPFIGKP 182

Query: 183 NAIIMNKSLEVLGIQRSEAVMVGDNLYLTDIMAGIQNDIATILVTTGFTPEEVPPLPIQP 242
 +IIM +++ VLG SE +MVGDN Y TDIMAGI + T+LV TG T+ E + +P
 Sbjct: 183 ESIIMEQAMRVLGTDVSETLMVGDN YATDIMAGINAGMDTLVHTGVTKREHMTDDMEKP 242

25 Query: 243 DHVLSLDEW 252
 H + SL EW
 Sbjct: 243 THAIDSLTEW 252

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 207/250 (82%), Positives = 227/250 (90%), Gaps = 1/250 (0%)

Query: 3 YKGYLIDLDGTIYKGSRIIPAGERFIERLQEKGIPYMLVTNNTRTPESVQEMLRG-FNV 61
 YKGYLIDLDGTIY+GK+RIPAGERFI+RLQE+GIPY+LVTTNNTRTPE VQ ML F+V
 35 Sbjct: 3 YKGYLIDLDGTIYQGKNRIPAGERFIKRLQERGIPYLLVTNNTRTPPEMVQSMANQPHV 62

Query: 62 ETPLETIYTATMATVDYMDNMNRGKTAYVIGEEGLKKAIAADAGYVEDTKNPAYVVVGLDW 121
 ET +ETIYTATMATVDYMDNMNRGKTAYVIGE GLK AIA AGYVE+ +NPAYVVVGLD
 Sbjct: 63 ETSIETIYTATMATVDYMDNMNRGKTAYVIGETGLKSAIAAAGYVEELENPAYVVVGLDS 122

40 Query: 122 NVTYDKLATATLAIQNGALFIGTNPDLNIPTERGLLPAGAGSLNALLEAATRIKPVFIGKP 181
 VTY+ LA ATLAIQ GALFIGTNPDLNIPTERGL+PGAG+LNALLEAATR+KPVFIGKP
 Sbjct: 123 QVTYEMLAIAATLAIQKGFALFIGTNPDLNIPTERGLMPGAGALNALLEAATRVKPVFIGKP 182

45 Query: 182 NAIIMNKALEILNIPRNQAVMVGDNLYLTDIMAGINNDIDTLLVTTGFTTVEEVPDLPIQP 241
 NAIIMNK+LE+L I R++AVMVGDNLYLTDIMAGI NDI T+LVTTGFT EEVP LPIQP
 Sbjct: 183 NAIIMNKSLEVLGIQRSEAVMVGDNLYLTDIMAGIQNDIATILVTTGFTPEEVPPLPIQP 242

Query: 242 SYVLASLDEW 251
 +VL+SLDEW
 50 Sbjct: 243 DHVLSLDEW 252

A similar DNA sequence was identified in *S.pyogenes* <SEQ ID 4215> which encodes amino acid sequence <SEQ ID 4216>. An alignment of the GAS and GBS sequences follows:

Identities = 94/204 (46%), Positives = 139/204 (68%)

55 Query: 4 VMVENTKLLCSWVWLLALAILITYSTWLWYPLEVDHLKLEQVVFMSKDAILHNYNGLIN 63
 +M + ++ +W+W+++LA L TIY WL YP+E+ LKLE+VV++ + I +N+N L+
 Sbjct: 5 LMKDKLLVLTWIIWISLATLATIYIAWLIYPIETQFLKLEKVVYLKAETIYYNFKLMI 64

60 Query: 64 YLTNPFVTRLEFANFHSSADGLKHFADVKWLFHLTQVVFGLLYPTLKTFTQRLKTKRFW 123
 YLT+PF++ L +F SS DGLKHFADVK+LF L +F+ L +P + + K K +
 Sbjct: 65 YLTHPFISDLNMPSPFSSDGLKHFADVKYLFHTLAHGLFVILTFPVIYFLRRGWKQKSIF 124

Query: 124 LLQKPLILALFPLMIGLMAFIGFEHFFTLFHQVLFVGDSSWLFDPDKDSVIWILPEVF 183

-1510-

L + +A + P+ I + A +GF+ FFTLFH+VLF GDS+W F+PL D VIWILPE F
 Sbjct: 125 LYEGFFKIAIMLPFIIVVCAFLGLGFDQFFTLFHEVLFPDSTWQFNPLTDPVIWILPETF 184
 Query: 184 FLHCFLFFMIVYEIILWSLVGLAR 207
 FLHCF+ F+++YE I L+ + R
 Sbjct: 185 FLHCFIIFLLIYETITITILLIIGR 208

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 1378

A DNA sequence (GBSx1463) was identified in *S.agalactiae* <SEQ ID 4221> which encodes the amino acid sequence <SEQ ID 4222>. This protein is predicted to be oleoyl-acyl carrier protein thioesterase. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3332(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB02069 GB:AB026647 acyl carrier protein thioesterase
 [Arabidopsis thaliana]
 Identities = 59/248 (23%), Positives = 104/248 (41%), Gaps = 30/248 (12%)
 Query: 2 GLLYRETYEVPFYESDTNHYMKLPQLLALALQISAKQSLKLGIGDD-----IVFKRYGLV 56
 GL Y+E + V YE +N + + L ++ + +G D ++ L+
 Sbjct: 81 GLSYKEKFVVRSEYVGSNKTATVETIANLLQEVGCNHAQSVGFSTDGFATTTTMRKLHLI 140
 Query: 57 WVVTDYIIDIERLPKHAEKIVITEAKAHNKLCCYRYFYIYGE-DGQKIITISSAFVFLMD 115
 WV I+I + P + + IET ++ ++ R + + G+ +S +V+M+
 Sbjct: 141 WVTARMHIEIYKPAWGDVVEIETWCQSEGRIGTRRDWILKDSVTGGEVTGRATSKWVMMN 200
 Query: 116 FKTRKIHPVLDDITSY-----QSQRIKKVIRGPKYHPIGDSKVKQYHVR 160
 TR++ V DD+ Y ++ +KK+ PK + R
 Sbjct: 201 QDTRRLQKVSDDVRDEYLVFCPEPRLAPFPEENRSLKKI---PKLEDPAQYSMIGLKPR 257
 Query: 161 YFDLDMNGHVNNNSKYLEWMYDVLDFLSSHIPKKIDLKYYKEIQYGTDIKSHWYQDGLV 220
 DLDMN HVNN Y+ W+ + + + +H + I L Y +E Q + D L
 Sbjct: 258 RADLDMNQHVNNVTYIGWVLESIPQEIVDTHELQVITLDYRRECQDDVV-----DSLTF 311
 Query: 221 TRHDIIGG 228
 T IGG
 Sbjct: 312 TTTSEIGG 319

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4223> which encodes the amino acid sequence <SEQ ID 4224>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.88 Transmembrane 21 - 37 (21 - 38)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2550(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-1511-

>GP:AAB71730 GB:U65643 acyl-ACP thioesterase [Myristica fragrans]
Identities = 41/128 (32%), Positives = 67/128 (52%), Gaps = 11/128 (8%)

5 Query: 33 FIFMIKRGGLVDILAYFALLNPDTRKVTATIPEDLVAPFETDFVKKLHRV-----PKMPL 87
F+ K G +L + + ++N TR+++ IPE++ E FV+ H V K+P
Sbjct: 147 FLRDCKTGELTRATSVVVMNKRTRRLSKIPEEVRVEIEPYFVE--HGVLDSDSRKLEK 204

10 Query: 88 LEQS----IDRDYVRYFDIDMNGHVNNISKYLDWMYDVLGCEFLKTHQPLKMTLKYYKEV 143
L + I R R+ D+D+N HVNN KY+ W+ + + L++H+ MTL+Y KE
Sbjct: 205 LNDNTANYIRGLAPRWSDLVNQHVNNVKYIGWILESVPSLLESHELYGMTLEYRKEC 264

15 Query: 144 SPGGQITS 151
G + S
Sbjct: 265 GKDGLLQS 272

An alignment of the GAS and GBS proteins is shown below.

Identities = 62/144 (43%), Positives = 94/144 (65%)

20 Query: 101 GOKIITISSAFVLMDFKTRKIHPVLDDITSYQSQRIRKQVIRGPKYHPIGDSKVKQYHVR 160
G ++ I + F L++ TRK+ + +D+ + +++ +KK+ R PK + S + Y+VR
Sbjct: 40 GGLLDILAYFALLNPDTRKVTATIPEDLVAPFETDFVKKLHRVPMPLLEQSIDRDYVVR 99

25 Query: 161 YFDLDMNGHVNNISKYLEWYDVLDDLFLSSHIPKKIDLKYYIKEIQYGTDIKSHWYQDGLV 220
YFD+DMNGHVNNISKYL+WMYDVL +FL +H P K+ LKY+KE+ G I S ++ D L
Sbjct: 100 YFDIDMNGHVNNISKYLDWMYDVLGCEFLKTHQPLKMTLKYYKEVSPGGQITSSYHLDQLT 159

30 Query: 221 TRHDIIGDAIHAQARIEWQEKKE 244
+ H I ++AQA IEW+ K+
Sbjct: 160 SYHQITSDGQLNAQAMIEWRAIKQ 183

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1379

35 A DNA sequence (GBSx1464) was identified in *S. agalactiae* <SEQ ID 4225> which encodes the amino acid sequence <SEQ ID 4226>. This protein is predicted to be coproporphyrinogen III oxidase. Analysis of this protein sequence reveals the following:

Possible site: 40
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1484(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05062 GB:AP001511 coproporphyrinogen III oxidase [Bacillus halodurans]
Identities = 173/375 (46%), Positives = 248/375 (66%), Gaps = 5/375 (1%)

50 Query: 5 PTSAYVHIPFCTQICYCDFSKVFIKNQPVDAYLQALIREFR-----SYDITELRTLYIGG 60
P +AY+HIPFC ICYYCDF+K ++KNQPV+ YLQAL E L+TLY+GG
Sbjct: 2 PKAAYIHIPFCEHICYCDFNKFYLNQPVNEYLQALETEMAMVVAEQPTKSLQTLTVGG 61

55 Query: 61 GTPTSISAVQLDYLLTELSRDLNLTLEEFTEANPGDLTVDKIEVLQSAVNRVSLGVQ 120
GTPT+++A QL LL + R L L+ LEEFT E NP + +K++VL+ V+R+S+GVQ
Sbjct: 62 GTPTALTADQLAQLLASIKRTLPLSDLEEFTEFVNPDSEDEKLDVLRSYGVDRLSIGVQ 121

60 Query: 121 TFNDKHLKRIGRSHNEAQIYSTIDALKTAGFQNISIDLIYALPGQTMDDVRSNVAKALS 180
F LK IGR+H++ + ++ + AGF N+S+DL+ LP QT + + +A +L
Sbjct: 122 AFQPLLLKEIGRTHDQKSVQAVEKSRQAGFANLSLDMGLGPKQTPEMFAETLKEAFAL 181

-1512-

Query: 181 NIPHLSLYSLILEHHTVFMNKMRRGKHLPTEDLEAEMFEYIIISEMERNGFEHYEISNFT 240
 + HLS YSL +E TVF N+ R+G+L LP ED E +M+ + E E++GF+ YEISNF
 Sbjct: 182 EVEHLSYSLKVEAKTVFYNRQGRLLTLPPEDEVKMYRQLCYETEKHGFKQYEISNFA 241

5 Query: 241 KPGFESRHNLMYWDNVEYYGVGAGASGYLDGIRYRNRPQIHYLKGVSSEGNARLSE-EVL 299
 K G+ESRHNLMYWDNVEYYG GAGA GY+ G+RY N GP+ YL+ + EG + E +
 Sbjct: 242 KKGYESRHNLMYWNDEYYGFGAGAHGYVGGVRYMNHGPLPKYLQAMEEGRRPVFESHV 301

10 Query: 300 SKNEMMEEELFLGLRKKEGVSIGKFEQKFGTSFEKRYGQIVQELQSDGLLKENNGFIQMT 359
 S+ E MEE++FLGLRK+ GV F ++FG S Y + + +L + LL+ + +++T
 Sbjct: 302 SRVEQMEEQMFLGLRKRSVGEERVFVERFGVSMFSLYEKQIAQLVARCLLERTDDRVRILT 361

Query: 360 KKGLFLGDTVAEKFI 374
 +GL LG+ V E+F+
 15 Sbjct: 362 DEGLLLGNEVFQFL 376

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4227> which encodes the amino acid sequence <SEQ ID 4228>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3202(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 304/376 (80%), Positives = 343/376 (90%)

30 Query: 1 MLKKPTSAYVHIPFCTQICYCDFSKVFIKNQPVDAYLQALIREFRSYDITELRTLYIGG 60
 M KKPTSAYVHIPFCTQICYCDFSKVFI+NQPVDAYL+ALI+EF SY I +L+TLYIGG
 Sbjct: 33 MSKKPTSAYVHIPFCTQICYCDFSKVFIQNQPVDAYLKALIQEFD SYGIRD LKTLTYIGG 92

35 Query: 61 GTPTSISAVQLDYLLTELSRDNLNLTLEEFTEIANPGDLTVDKIEVLQSAVNRVSLGVQ 120
 GTPT+I+A QL+YLL L R+INL+ LEEFTIEANPGDLT +KI VLQ+SAVNR+SLGVQ
 Sbjct: 93 GTPTAITAKQLEYLLNHLERNLNLDDLEEFTEIANPGDLTPEKIAVLQRSVNRISLGVQ 152

40 Query: 121 TFNDKHLKRIGRSHNEAQIYSTIDALKTAGFQNISIDLIYALPGQTMDDVRSNVAKALS 180
 TFN+K LK+IGRSHNE QIYSTI LKTAGE NISIDLIYALPGQT+D V+ NVAKAL+L
 Sbjct: 153 TFNNKQLKQIGRSHNEEQIYSTIANLKTAGFHNISIDLIYALPGQTLQVKNVAKALAL 212

45 Query: 181 NIPHLSLYSLILEHHTVFMNKMRRGKHLPTEDLEAEMFEYIIISEMERNGFEHYEISNFT 240
 +IPHLSLYSLILEHHTVFMNKMRRGK+LPTEDLEAEMFEYIIISEME NGFEHYEISNFT
 Sbjct: 213 DIPHLSLYSLILEHHTVFMNKMRRGKLNLTEDLEAEMFEYIIISEMEANGFEHYEISNFT 272

Query: 241 KPGFESRHNLMYWDNVEYYGVGAGASGYLDGIRYRNRPQIHYLKGVSSEGNARLSEEVLS 300
 KPGFESRHNLMYWDNVEY+G GAGASGYL+GIRY+NR PIQHYLK V GNARL+EEVL
 Sbjct: 273 KPGFESRHNLMYWDNVEYFGCGAGASGYLNGIRYQNRVPIQHYLKAVEAGNARLNBEVLR 332

50 Query: 301 KNEMMEEELFLGLRKKEGVSIGKFEQKFGTSFEKRYGQIVQELQSDGLLKENNGFIQMTK 360
 K EMMEELFLGLRK GVS I +F++KFG SFE+RYG IV+ELQ+ GLL +++ F++MTK
 Sbjct: 333 KEEMMEEELFLGLRKKTGVS IQRFQKFGMSFEERYGNIVRELQNQGLLVKDDAFVRMTK 392

55 Query: 361 KGLFLGDTVAEKFIVE 376
 KGLFLGD+VAE+FI++
 Sbjct: 393 KGLFLGDSVAERFILD 408

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1513-

Example 1380

A DNA sequence (GBSx1465) was identified in *S.agalactiae* <SEQ ID 4229> which encodes the amino acid sequence <SEQ ID 4230>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3729(Affirmative) < succ>
10      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1381

A DNA sequence (GBSx1466) was identified in *S.agalactiae* <SEQ ID 4231> which encodes the amino acid sequence <SEQ ID 4232>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2989(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4233> which encodes the amino acid sequence <SEQ ID 4234>. Analysis of this protein sequence reveals the following:

```

Possible site: 57
30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2993(Affirmative) < succ>
35      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 36/109 (33%), Positives = 58/109 (53%), Gaps = 6/109 (5%)
40
Query: 9  WAKHKYLVLSKSKQKIYLDIRQTLKSPNCT---VLDVQSLIDQAVLLEESPSQVINYMH 65
      WA  KY V++ SQ+ Y +R+ K   + VL  LI++A + + + AY H+
Sbjct: 13 WAYQKYVWMAHSQQHYNALRELFKGNQWSEEKVLTFHCLIEEAQAIPPTVKSLRTAYQHV 72

Query: 66 WGYFKNKAERQEKEEFLLLEKYRKTGYQRRKLLAFLKQLLAKYPNSYL 114
      WGYFK A ++EK+ F L +   + ++L FL+++ A Y SYL
Sbjct: 73 WGYFKKVASQEEKDHFKDLDQALET---KSEEMLCFLQEMTAHYQPSYL 118

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1514-

Example 1382

A DNA sequence (GBSx1467) was identified in *S.agalactiae* <SEQ ID 4235> which encodes the amino acid sequence <SEQ ID 4236>. This protein is predicted to be mrsA (mrsA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 35
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.96    Transmembrane    56 - 72 ( 56 - 72)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAB11970 GB:Z99105 similar to phosphoglucomutase (glycolysis)
    [Bacillus subtilis]
    Identities = 284/451 (62%), Positives = 353/451 (77%), Gaps = 4/451 (0%)

20  Query: 1  MGKYFGTDGVRGEANVELTPELAFKLRFGGYVLSQHETDRPRVFVARDTRISGEMLESA 60
      MGKYFGTDGVRG AN ELTPELAFK+GRFGGYVL++ + RP+V + RDTRISG MLE A
      Sbjct: 1  MGKYFGTDGVRGVANSELTPELAFKVGGRFGGYVLTK-DKQRPKVLIGRDTRISGHMLEGA 59

      Query: 61  LIAGLLSVGIEVYKLGVLATPGVSYLVRTEKASAGVMISASHNPALDNGIKFFGSDGFKL 120
      L+AGLLS+G EV +LGV++TPGVSYL + A AGVMISASHNP DNGIKFFG DGFKL
25  Sbjct: 60  LVAGLLSIGAEVMRLGVISTPGVSYLTKAMDAEAGVMISASHNPVQDNGIKFFGGDGFKL 119

      Query: 121  DDDRELEIEALLDAKEDTLPRPSAQGLGLTLDVYPEGLRKYEFKFMESTGI-DLEGMKVALD 179
      D++E EIE L+D ED LPRP LG + DY EG +KY +F++ T D G+ VALD
30  Sbjct: 120  SDEQEAEIERIMDEPEDKLRPVGADLGLVNDYFEGGQKYLQFLKQTADDEFITGIHVALD 179

      Query: 180  TANGAATASARNIFLDLNADISVIGDQPDGLNINDGVGSTHPEQLQSLVRENGSDIGLAF 239
      ANGA ++ A ++F DL+AD+S +G P+GLNINDGVGSTHPE L + V+E +D+GLAF
      Sbjct: 180  CANGATSSLATHLFADLDADVSTMGTSPNGLNINDGVGSTHPEALSAFVKEKNADLGLAF 239

35  Query: 240  DGDSDRLIAVDENGEIVDGDKIMFIIGKYLSDKGQLAQNITVTTVMSNLGFHKALDREGI 299
      DGD DRLIAVDE G IVDGD+IM+I K+L +G+L +T+V+TVMSNLGF+KAL++EGI
      Sbjct: 240  DGDGDRLIAVDEKGNIVDGDQIMYICKHLKSEGRKDDTVVSTVMSNLGFYKALEKEGI 299

40  Query: 300  HKAITAVGDRYVVEEMRKSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLTKVMKETGKKL 359
      TAVGDRYVVE M+K GYN+GGEQSGH+I +DYNTTGDG L+AI L +K TGK L
      Sbjct: 300  KSVQTAVGDRYVVEAMKKGYNVGGEQSGHLIFLDYNTTGDGLLSAIIMLMNTLKATGKPL 359

      Query: 360  SELASEVTIYPQKLVNIRVENNMKDKAMEVPAIAEIIAKMEEEMDNGRILVRPSGTEPL 419
      SELA+E+ +PQ LVN+RV + K K E + +I+++E+EM+G+GRILVRPSGTEPL
45  Sbjct: 360  SELAAEMQKFPQLLVNVRVTD--KYKVEENEKVKAVISEVEKEMNGDGRILVRPSGTEPL 417

      Query: 420  LRVMAEAPTNEAVDYVDTIADVVRTEIGLD 450
      +RVMAEA T E D YV+ I +VVR+E+GL+
50  Sbjct: 418  VRVMAEAKTKELCDEYVNRIVEVVRSEMGLE 448

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4237> which encodes the amino acid sequence <SEQ ID 4238>. Analysis of this protein sequence reveals the following:

```

55  Possible site: 35
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.96    Transmembrane    56 - 72 ( 56 - 72)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
60  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-1515-

The protein has homology with the following sequences in the databases:

>GP:CAB11970 GB:Z99105 similar to phosphoglucomutase (glycolysis)
[Bacillus subtilis]

Identities = 287/451 (63%), Positives = 346/451 (76%), Gaps = 4/451 (0%)

5 Query: 1 MGKYFGTDGVRGEANVELTPELAFKLGFRGGYVLSQHETERPKVFVARDTRISGEMLESA 60
MGKYFGTDGVRG AN ELTPELAFK+GRFGGYVL++ + +RPKV + RDTRISG MLE A
Sbjct: 1 MGKYFGTDGVRGVANSELTPELAFKVGFRGGYVLTK-DKQRPKVLIGRDTRISGHMLEGA 59

10 Query: 61 LIAGLLSVGIEVYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFFGNDGFKL 120
L+AGLLS+G EV +LGV++TPGVSYL + A AGVMISASHNP DNGIKFFG DGFKL
Sbjct: 60 LVAGLLSIGAEVMRLGVISTPGVSYLTKAMDAEAGVMISASHNPVQDNGIKFFGCGDGFKL 119

15 Query: 121 ADDQELEIEALLDAPEDTLPRPSAEGGLTLDVYPEGLRKYKFLVTTGT-DLSGMTVALD 179
+D+QE EIE L+D PED LPRP LG + DY EG +KY +FL T D +G+ VALD
Sbjct: 120 SDEQEAEIERLMDEPEDKLPREFVGADLGLVNDYFEGGQKYLQFLKQTADDEFTGIHVALD 179

20 Query: 180 TANGAASVSARDVFLDLNAEIAVIGEKPNGLNINDGVGSTRPEQLQELVKETGADLGLAF 239
ANGA S A +F DL+A+++ +G PNGLNINDGVGST PE L VKE ADLGLAF
Sbjct: 180 CANGATSSLATHLFADLDADVSTMGTSPNGLNINDGVGSTHPEALSAFVKEKNADLGLAF 239

25 Query: 240 DGSDRLIAVDDETGEIVDGDRIIMFIIGKYLSEKGLLAHNTIVTTVMNLGFHKALDKQGI 299
DGD DRLIAVDE G IVDGD+IM+I K+L +G L +T+V+TVMSNLGF+KAL+K+GI
Sbjct: 240 DGDGRLIAVDEKGNIVDGDQIMYICSHLSEKGLKDDTVVSTVMNLGFYKALEKEGI 299

30 Query: 300 NKAITAVGDRYVVEEMRSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLAKVMKETGKSL 359
TAVGDRYVVE M+ GYN+GGEQSGH+I +DYNTTGDG L+AI L +K TKG L
Sbjct: 300 KSVQTAVGDRYVVEAMKKGYNVVGGEQSGHLIFLDYNTTGDGLLSAIMLMNTLKATGKPL 359

35 Query: 360 SELAAEVTIYPQKLVNIRVENSMMKERAMEVPAIANIIAKMEDEMAGNGRILVRPSGTEPL 419
SELAAE+ +PQ LVN+RV + K + E + +I+++E EM G+GRILVRPSGTEPL
Sbjct: 360 SELAAEMQKFPQQLLVNVRVTD--KYKVEENEKVKAVISEVEKEMNCGRILVRPSGTEPL 417

40 Query: 420 LRVMAEAPTDAEVDYVDTIADVVRTEIGCD 450
+RVMAEA T D YV+ I +VVR+E+G +
Sbjct: 418 VRVMAEAKTKELCDEYVNRIVEVVRSEMGLE 448

An alignment of the GAS and GBS proteins is shown below.

Identities = 400/450 (88%), Positives = 429/450 (94%)

40 Query: 1 MGKYFGTDGVRGEANVELTPELAFKLGFRGGYVLSQHETDRPRVFVARDTRISGEMLESA 60
MGKYFGTDGVRGEANVELTPELAFKLGFRGGYVLSQHET+RP+VFVARDTRISGEMLESA
Sbjct: 1 MGKYFGTDGVRGEANVELTPELAFKLGFRGGYVLSQHETERPKVFVARDTRISGEMLESA 60

45 Query: 61 LIAGLLSVGIEVYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFFGSDGFKL 120
LIAGLLSVGIEVYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFFG+DGFKL
Sbjct: 61 LIAGLLSVGIEVYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFFGNDGFKL 120

50 Query: 121 DDDRELEIEALLDAKEDTLPRPSAQLGLTLDVYPEGLRKYKFMESTGIDLEGMKVALDT 180
DD+ELEIEALLDA EDTLPRPSA+GLGLTLDVYPEGLRKYKFL+ +TG DL GM VALDT
Sbjct: 121 ADDQELEIEALLDAPEDTLPRPSAEGGLTLDVYPEGLRKYKFLVTTGTDLSGMTVALDT 180

55 Query: 181 ANGAATASARNIFLDLNADISVIGDQPDGLNINDGVGSTRPEQLQSLVRENGSDIGLAFD 240
ANGAA+ SAR++FLDLNA+I+VIG++P+GLNINDGVGST PEQLQ LV+E G+D+GLAFD
Sbjct: 181 ANGAASVSARDVFLDLNAEIAVIGEKPNGLNINDGVGSTRPEQLQELVKETGADLGLAFD 240

60 Query: 241 GDSDRLIAVDENGEIVDGDKIMFIIGKYLSDKGQLAQNITIVTTVMNLGFHKALDREGIH 300
GDSDRLIAVDE GEIVDGD+IMFIIGKYL+KG LA NTIVTTVMNLGFHKALD++GI+
Sbjct: 241 GDSDRLIAVDDETGEIVDGDRIIMFIIGKYLSEKGLLAHNTIVTTVMNLGFHKALDKQGIN 300

65 Query: 301 KAITAVGDRYVVEEMRSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLTKVMKETGKKLS 360
KAITAVGDRYVVEEMR SGYNLGGEQSGHVIIMDYNTTGDGQLTAIQL KVMKETGK LS
Sbjct: 301 KAITAVGDRYVVEEMRSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLAKVMKETGKLS 360

70 Query: 361 ELASEVTIYPQKLVNIRVENNMKDKAMEVPAIAEIIAKMEEMDGNRILVRPSGTEPLL 420
ELA+EVTTIYPQKLVNIRVEN+MK++AMEVPAIA IIAKME+EM GNGRILVRPSGTEPLL

-1516-

Sbjct: 361 ELAAEVTIYPQKLVNIRVENSMKERAMEVPAIANIIAKMEDEMAGNGRILVRPSGTEPLL 420

Query: 421 RVMAEAPTNEAVDYYVDTIADVVRTEIGLD 450

RVMAEAPT+ VDYYVDTIADVVRTEIG D

5 Sbjct: 421 RVMAEAPTDAEVDYYVDTIADVVRTEIGCD 450

SEQ ID 4236 (GBS402) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 5; MW 78kDa).

GBS402-GST was purified as shown in Figure 218, lane 3-5.

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1383

A DNA sequence (GBSx1468) was identified in *S. agalactiae* <SEQ ID 4239> which encodes the amino acid sequence <SEQ ID 4240>. Analysis of this protein sequence reveals the following:

15 Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

20 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:CAB11969 GB:Z99105 ybbR [Bacillus subtilis]

Identities = 90/324 (27%), Positives = 167/324 (50%), Gaps = 18/324 (5%)

Query: 1 MKKFFTNKFWLGVVSLFLAILLFLTATATSMNHQDNSKIAG-----ASETYHTLTLDVPI 55

M KF N++ + +++L A+LL++ A + N K G S T TLTD+P+

30 Sbjct: 1 MDKFLNNRWAVKIIALLFALLLYV---AVNSNQAPTPKKPGESFFPTSTTDEATLTIDIPV 57

Query: 56 DIKYDSDDYFISGYSGADVMS-SVNRVKLDSEINEDTRKFQVADLTNMPGTHKVP 114

YD ++Y ++G +V + S + VK + T+ F++ AD+ ++K GTHKV L

Sbjct: 58 KAYYDDENYVVTGVPQTVNVTIKGSTSAVKARQ----TKNFEIYADMEHLKTGTHKVEL 113

35 Query: 115 KVVNLPSPGVNATVSPTTITVTMGKKKTKEFPV-YGHVNDKQIKAGYAVDKMSVDVSKVKV 173

K N+ G+ +++P+ TVT+ ++ TK FFPV + N ++K GY+ ++ V V++

Sbjct: 114 KAKNVSDGLTISINPSVTTVTIQERTTKSFPVEVEYYNKSMMKKCYSPQPIVSPKNVQI 173

40 Query: 174 TSDESIIDRIDHVAANIPDDKVLDDDFNKTIVTLQAVTADGTVIASIIHPSKATLSVKVKK 233

T +++ID I A++ + D+ K + DG L + PS ++V V

Sbjct: 174 TGSKNVIDNISLHKASVNLNA-DETIEKEAKVTYVDKGNALPVDVEPSVIKITVPVTS 232

Query: 234 LTKTVPINLIPVGQFSDSISKINYKLSQEKAVISGTKEALEAISVIN-AEVDISDVTKNT 292

+K VP + G D+S N + S + + G+++ L+++ I+ +D+S + K++

45 Sbjct: 233 PSKKVPFKIERTGSLPDGVSIANIESSPSEVTYGSQDVLDSLEFIDGVSLDLKINKDS 292

Query: 293 --EKINLSANNVSVDPAQVTVQL 314

E I L + P++VT+ +

50 Sbjct: 293 DIEADIPLPDGVKKISPSKVTLHI 316

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4241> which encodes the amino acid sequence <SEQ ID 4242>. Analysis of this protein sequence reveals the following:

Possible site: 29

55 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

-1517-

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP:CAB11969 GB:Z99105 ybbR [Bacillus subtilis]

Identities = 81/322 (25%), Positives = 154/322 (47%), Gaps = 15/322 (4%)

10 Query: 1 MKRFLNSRPWLGMVSVFFAILLFLTAASSNH---NNSSSQIYSPIETYTHSLKDVPI DM 56
 M +FLN+R + ++++ FA+LL++ A +SN + T +L D+P+
 Sbjct: 1 MDKFLNNRWAVKIIALLFALLLYV-AVNSNQAPTPKKPGESFFPTSTTDEATLTIDIPVKA 59

15 Query: 57 KYDSDKYFISGYSYGAEVYLT-STNRIKLDSEVNNDTRNFKIVADLTHSHPGTVSVNLRV 115
 YD + Y ++G V + ST+ +K + T+NF+I AD+ H GT V L+
 Sbjct: 60 YYDDENYVVTGVPTVNTIKGSTSAVKKARQ---TKNFEIYADMEHLKTGTHKVELKA 115

20 Query: 116 ENLPSGVTATVSPDKISVTIGKKESKVFPVRGS-VDAXQIANGYEISKIETGVNKVEVTS 174
 +N+ G+T +++P +VTI ++ +K FPV + ++ GY + V++T
 Sbjct: 116 KNVSDGLTISINPSVTTVTIQERTTSFPPVEVEYYNKSKMKKGYSPQPIVSEKPNVQITG 175

25 Query: 175 DESTIALIDHVVAKLPDDQVLDNRNYSRVTLQAVSADGTILASAIIDPAKTNLSVAVKKIT 234
 ++ I I A + + D + DG L ++P+ ++V V +
 Sbjct: 176 SKNVIDNISLHKASVNLNA-DETIEKEAKVTVDKGNALPVDVEPSVIKITVPTVSPS 234

30 Query: 235 KSVPIRVEAVGMDDSLSDIQYKLSKQTAVISGSREVLEDIDEII-AEVDNISDVTKNT-- 291
 K VP ++E G + D +S + S + GS++VL+ ++ I +++S + K++
 Sbjct: 235 KKVPFKIERTGSLPDGVSIANIESSPSEVTVYGSQDVLDSLEFIDGVSLDLSKINKDSDI 294

Query: 292 SKTVSLSSSQVSIIEPSVVTVQL 313
 + L I PS VT+ +
 Sbjct: 295 EADIPLPDGVKKISPSKVILHI 316

An alignment of the GAS and GBS proteins is shown below.

Identities = 198/319 (62%), Positives = 251/319 (78%), Gaps = 1/319 (0%)

35 Query: 1 MKKFFTNKFWLGVVSLFLAILLFLTATATSMNHQDNSKIAGASETYHTLTIDVPIDIKYD 60
 MK+F ++ WL+VS+F AILLFLTA A+S ++ +S+I ETYTH+L DVPID+KYD
 Sbjct: 1 MKRFLNSRPWLGMVSVFFAILLFLTA-ASSNHNSSSQIYSPIETYTHSLKDVPI DMKYD 59

40 Query: 61 SDDYFISGYSYGADVYMSSVNRVKLDSEINEDTRKFKVADLTNMKFGTHKVLKVVNLP 120
 SD YFISGYSYGA+VY++S NR+KLDSE+N DTR FK+VADLT+ PGT V L+V NLP
 Sbjct: 60 SDKYFISGYSYGAEVYLTSTNRIKLDSEVNNDTRNFKIVADLTHSHPGTVSVNLRVENLP 119

45 Query: 121 SGVNATVSPTTITVTMGKKKTKEFPVYGHVNDKQIKAGYAVDKMSVDVSKVKVTSDESII 180
 SGV ATVSP I+VT+GKK++K FPV G V+ KQI GY + K+ V+KV+VTSDES I
 Sbjct: 120 SGVTATVSPDKISVTIGKKESKVFPVRGSVDAXQIANGYEISKIETGVNKVEVTSDESTI 179

50 Query: 181 DRIDHVAANIPDDKVLDDDFNKTVTTLQAVTADGTVLASIIHPSKATLSVKVKKLTKTVP I 240
 IDHV A +PDD+VLD +++ VTLQAV+ADGT+LAS I P+K LSV VKK+TK+VPI
 Sbjct: 180 ALIDHVVAKLPDDQVLDNRNYSRVTLQAVSADGTILASAIIDPAKTNLSVAVKKITKSVPI 239

55 Query: 241 NLIPVGQFSDSISKINYKLSQEKAVISGTKEALEAISVINAEDISDVTKNTEKKINLSA 300
 + VG DS+S I YKLS++ AVISG++E LE I I AEV+ISDVTKNT K ++LS+
 Sbjct: 240 RVEAVGMDDSLSDIQYKLSKQTAVISGSREVLEDIDEIIAEVDNISDVTKNTSKTVSLSS 299

Query: 301 NNVSVDPAQVTVQLTTTKK 319
 + VS++P+ VTVQLTTTKK
 Sbjct: 300 SQVSIIEPSVVTVQLTTTKK 318

60 SEQ ID 4240 (GBS99) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 6; MW 35.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 9; MW 60.7kDa).

-1518-

The GBS99-GST fusion product was purified (Figure 197, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 293), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1384

A DNA sequence (GBSx1469) was identified in *S.agalactiae* <SEQ ID 4243> which encodes the amino acid sequence <SEQ ID 4244>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0503(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1385

A DNA sequence (GBSx1470) was identified in *S.agalactiae* <SEQ ID 4245> which encodes the amino acid sequence <SEQ ID 4246>. Analysis of this protein sequence reveals the following:

Possible site: 15
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.50	Transmembrane	20 - 36 (18 - 46)
INTEGRAL	Likelihood = -7.64	Transmembrane	48 - 64 (42 - 68)
INTEGRAL	Likelihood = -3.40	Transmembrane	80 - 96 (80 - 96)

----- Final Results -----

bacterial membrane --- Certainty=0.4800(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11968 GB:Z99105 alternate gene name: ybbQ-similar to
hypothetical proteins [Bacillus subtilis]
Identities = 125/253 (49%), Positives = 186/253 (73%), Gaps = 5/253 (1%)

Query: 27 MDIIIVAVLIYKFIKALAGTKIMSLIQSVILFIIIRFVSEWIGLTTITFLMNQIVTYGVI 86
+DI++V +IYK I + GTK + L++G+++ ++R S+++GL+T+ +LM+Q +T+G +
Sbjct: 16 VDILLVWYVIYKLIMVIRGTRKAVQLLKGIVVIVLVRMASQYLGLSTLQWLMDQAITWGFL 75

Query: 87 AGVVIFAPEIRITGLEKFGRTPLQFTQRSQLSSDE---KLVDALVKAVAYMSPRKIGALIS 143
A ++IF PE+R LE+ GR F RS +E K ++A+ KA+ YM+ R+IGAL++
Sbjct: 76 AIIILFQPELRRALEQLGRGR--FFSRSGTFVEEAQQKTIEAITKAINYMAKRRIGALIT 133

Query: 144 IERTQTLQEIYIATGIPLDADISSELLINIFIPNTPLHDGAVIVKDKKIATACSYLPLSES 203
IER + +YI TGIPL+A +SSELLINIFIPNTPLHDGAVI+K+ +IA A YLPLSES
Sbjct: 134 IERDTGMGDYIETGIPLNAKVSSELLINIFIPNTPLHDGAVIMKNEIAAACYLPLSES 193

Query: 204 SSISKEFGTRHRAAIGLSSENSDALTVIVSEETGGISVALKGEFLHDLKDSFEAILRTQL 263

-1519-

ISKE GTRHRAA+G+SE +D+LT+IVSEETGG+SVA G+ +L++++ + +L +
 Sbjct: 194 PFISKELGTRHRAAVGISEVTDLSLTIIIVSEETGCVSVAKNGDLHRELTEELKEMLEAEF 253

Query: 264 IQNQEENSKLAWY 276

+N + S WY

Sbjct: 254 KKNTRDTSSNRWY 266

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4247> which encodes the amino acid sequence <SEQ ID 4248>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -6.64 Transmembrane 20 - 36 (19 - 40)
 INTEGRAL Likelihood = -6.21 Transmembrane 48 - 64 (47 - 68)
 INTEGRAL Likelihood = -2.07 Transmembrane 76 - 92 (76 - 92)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3654(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA03984 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 117/255 (45%), Positives = 178/255 (68%), Gaps = 6/255 (2%)

Query: 19 PWL-LAVHLLDILIVAYLIYRFIKALTGTKIMSLVQGVIFFLVLRFAEWIGFTTITYLM 77
 PWL +LDIL+V Y+IY+ I + GT+ + L++G+ L++ I+ + T+ +++
 Sbjct: 8 PWNLYLTQILDILVVTYVIYKAIMIIRGTTRAVQLLKGITVILIVYALSIFFNLRITLGVIV 67

Query: 78 NQVITYGVIAGVVIFTPEIRAGLEKFGRTQVFLQKQYVSSSEAL---VDALIKSVAYMG 134
 NQ IY+G++A ++IF PE+R LE+ GR F + + E + +DA++K+ YMG
 Sbjct: 68 NQAITYGLLAVIIIFQPELRRALQLGRGR--FFASRTANEETMKKTIDAIVKASTYMG 125

Query: 135 PRKIGALIAIEQTQTLQEYIATGIPLNADISSQLLINIFIPNTPLHDGAVIVGQNKIVAA 194
 R+IGALI++E+ + +Y+ TGIP+NA+++S+LLIN FIPNTPLHDGAVI+ + I+AA
 Sbjct: 126 KRRIGALISMERETGMDTYVETGIPMNANLTSELLINTFIPNTPLHDGAVIINNDTILAA 185

Query: 195 CAYLPLSESKAISKEFGTRHRAAIGLSENSDALTIIIVSEETGAISVTRKGQFLHDLSTDE 254
 YLPLSE+ ISKE GTRHRAA+G+SE +D LTI+VSEETG IS+T+ G+ DL ++
 Sbjct: 186 ACYLPLSENPFISKELGTRHRAALGVSEVIDCLTIVVSEETGHISLTKNGELHRDLDEEQ 245

Query: 255 FETVLRITYLMSNSNV 269
 ++L L+S + +
 Sbjct: 246 LRSLLEAELISEAKM 260

An alignment of the GAS and GBS proteins is shown below.

Identities = 201/283 (71%), Positives = 239/283 (84%), Gaps = 2/283 (0%)

Query: 1 MDIFSIDSKFASIMENPWWILIHLMDDIIIVAVLIYKFIKALAGTKIMSLIQGVILFII 60
 M+ S+ID KF S+ +PW++ +HL+DI+IVA LIY+FIKAL GTKIMSL+QGVIF++
 Sbjct: 1 MNNLSSIDIKFLLSLFADFPWLLAVHLLDILIVAYLIYRFIKALTGTKIMSLVQGVIFFLV 60

Query: 61 IRFVSEWIGLTTITFLMNQIVTYGVIAGVVIFAPEIRTGLEKFGRTQQLFTQSQLSSDE 120
 +RF++EWIG TTIT+LMNQ++TYGVIAGVVIF PEIR GLEKFG+ Q+F Q+ +SS+
 Sbjct: 61 LRFIAEWIGFTTITYLMNQVITYGVIAGVVIFTPEIRAGLEKFGRTQVFLQKQYVSSSE 120

Query: 121 KLVDALVKAVAYMSPRKIGALISIERQTTLQEYIATGIPLDADISSELLINIFIPNTPLH 180
 LVDAL+K+VAYM PRKIGALI+IE+TQTLQEYIATGIPL+ADISS+LLINIFIPNTPLH
 Sbjct: 121 ALVDALIKSVAYMGPRKIGALIAIEQTQTLQEYIATGIPLNADISSQLLINIFIPNTPLH 180

Query: 181 DGAVIVKDKKIATACSYLPLSESSSISKEFGTRHRAAIGLSENSDALTIVVSEETGGISV 240
 DGAVIV KI AC+YLPLSES +ISKEFGTRHRAAIGLSENSDALT+IVSEETG ISV
 Sbjct: 181 DGAVIVGQNKIVAACAYLPLSESKAISKEFGTRHRAAIGLSENSDALTIIIVSEETGAISV 240

Query: 241 ALKGEFLHDLSDSKSFEAILRTQLIQNQEENSKLAWYNQLLRK 283

-1520-

KG+FLHDL S D FE +LRT L+ N N L WY ++L K
 Sbjct: 241 TRKGQFLHDLSTDEFETVLRITYLMSN--SNVTLWPYKKILGGK 281

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1386

A DNA sequence (GBSx1471) was identified in *S.agalactiae* <SEQ ID 4249> which encodes the amino acid sequence <SEQ ID 4250>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.60 Transmembrane 33 - 49 (33 - 49)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1387

A DNA sequence (GBSx1472) was identified in *S.agalactiae* <SEQ ID 4251> which encodes the amino acid sequence <SEQ ID 4252>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1001(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9781> which encodes amino acid sequence <SEQ ID 9782> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC84012 GB:AF080002 UDP-N-acetylmuramyl tripeptide synthetase
 MurC [Heliobacillus mobilis]
 Identities = 143/442 (32%), Positives = 229/442 (51%), Gaps = 17/442 (3%)
 Query: 12 GKSAHYLLSKMGRGST-YPGSLALKFDKDILDTIAKDYE--IVVVGITNGKTLTALTIVG 68
 GK+A +L + G G T +PG + + IL +A+ + +VVIGTNGKT T+ +
 Sbjct: 2 GKTAIWLNRFRFGGTSFPGGIGRRVAPQILTALARQLKRGAMVVVGITNGKTTTSKMLAA 61
 Query: 69 ILKEAFGQVVTNPSGANMITGIVSTFLTAKSKSG--KKIAVLEIDEASLPRTQYIKPS 126
 I++++ + N +GAN++ GI + F+ + + ++E+DEA++P++ + ++P
 Sbjct: 62 IVEKSSLTLTNHRAGANLVGGITTAFIDSATIGGSIITSDLGIIIEVDEATIPQLVREVQPK 121
 Query: 127 LFFVTNIFRDQMDRYGEIYTTYQMILDGAANAP-QATILANGDSPLFNS--KSVTNPVQF 183
 V TN FRDQ+DR+GE+ T ++ + P Q+ + N D PL S K V +
 Sbjct: 122 GVVVTNFFRDQLDRFGELDKTVSLVGEALRLLPVQSI AVLNADDP LVASLGKDFPGRVLY 181
 Query: 184 YGFNTDKHEPRLAHYNTGILCPKCQAILTYRLNTYANLGDYTCPNCDFERP NLDYALTR 243
 +G + + R + E C C LTY + LG Y C +C FERP +T

-1521-

Sbjct: 182 FGIDDRSYGAREMLQSAETRFCRLCGHPLTYDWFFFGQLGHYRCSHCGFERPEPKIKVTG 241

Query: 244 LTHLTNTSSGFVIDGQ----QYNINVGGLYNIYNALAAVSAEYFGVEPSQIKDGFDKSR 299
 + S F ++ Q ++ G YNIYNALAA++ A + I+ G R

5 Sbjct: 242 IQLKGEGSFAFTVETPRGTWQLELSTPGFYNIYNALAAIASAIRLDLPEKAIRAGLQGYR 301

Query: 300 AVFGRQETFTIGN-KKCTLVLIKNPVGASQALDMIKLAPYPFSLSVLLNANYADGIDTSW 358
 FGR E ++ ++ L LIKNP G + + + P L V++N N ADG D SW

10 Sbjct: 302 TNFGRMERIELEDGRRRAFLALIKNPTGCEVIRTLVQNRGPKRLLVIINDNADGRDISW 361

Query: 359 IWDANFETI--LTMNIPFIFAGGVRHSEIARRLRVTGYDEKRIK-QADKLQDIMTMIEQQ 415
 +WDA+FE++ + + +F G+R ++A RL TG + I+ +A+ I + +E

Sbjct: 362 LWDADFESLEPVPELRSVFTSGLRGEDMALRLNYTGIPAESIRYEANVESAIRSALEMT 421

15 Query: 416 ET-EHAYILATYTMLEFREIL 436
 E E YIL TYTA+LE + L

Sbjct: 422 EPGETLYILEPTYTALLESKAAL 443

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4253> which encodes the amino acid sequence <SEQ ID 4254>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have an uncleavable N-term signal seq

25 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 343/446 (76%), Positives = 393/446 (87%)

Query: 1 MKINTALGVAAGKSAHYLLSKMRGSGTYPGSLALKFDKIDILDTIAKDYETVVVTGTNGKT 60
 MK+ T LG+ AGK+A +L+K+GRGSTYPG LAL DKDIL ++KDY+IVVVTGTNGKT

35 Sbjct: 1 MKMKITLLGIIAGKAAQSILTKLGRGSGTYPGRLALACDKDILKDLSDYDIVVVTGTNGKT 60

Query: 61 LTTALTIVGILKEAFGQVVTNPSGANMITGIVSTFLTAKKSKSGKKI AVL EIDEASLPRT 120
 LTTALTIVGILKEAFG+++TNPSGANMITGI STFL AKK KS ++I AVL EIDEASLPRT

Sbjct: 61 LTTALTIVGILKEAFGEIITNPSGANMITGITSTFLAAKKGKSERQI AVL EIDEASLPRT 120

40 Query: 121 QYIKPSLFVFTNIFRDQMDRYGEIYTTYQMILDGAANAPQATILANGDSPFNKSVTNP 180
 Y+KPSLFV+TNIFRDQMDRYGEIYTTYQMI+DGA NAP+ATILANGDSP+F+SK + NP

Sbjct: 121 TYLKPSLFVYTNIFRDQMDRYGEIYTTYQMIVDGARNAPKATILANGDSPIFSSKDIVNP 180

45 Query: 181 VQFYGFNTDKHEPRLAHYNTGILCPKCQAILTYRLNTYANLGDYTCPNCDFERPNLDYA 240
 VQ+YGF+T KH P+LAHYNTGILCPKC+ IL YRLNTYANLGD+ C NC F+RP LDY

Sbjct: 181 VQYVGFD TAKHAPQLAHYNTGILCPKCEHILQYRLNTYANLGDVCLNCQFQRPTLDYQ 240

Query: 241 LTRLTHLTNTSSGFVIDGQQYNINVGGLYNIYNALAAVSAEYFGVEPSQIKDGFDKSRA 300
 LT LT +T+ SS FVIDGQ Y INVGGLYNIYNALAAVSAE+FGV P +IK GF+KS+A

50 Sbjct: 241 LTELTAITHQSSEFVIDGQYKINVGGLYNIYNALAAVSAEFGVSPEKIKAGFNKSKA 300

Query: 301 VFGRQETFTIGNKKCTLVLIKNPVGASQALDMIKLAPYPFSLSVLLNANYADGIDTSWIW 360
 VFGRQETFT+G+K CTL+LIKNPVGASQAL+MI+LA YPFSLSVLLNANYADGIDTSWIW

55 Sbjct: 301 VFGRQETFTVGDKSCTLILIKNPVGASQALEMIQLADYPFSLSVLLNANYADGIDTSWIW 360

Query: 361 DANFETILT MNIPFIFAGGVRHSEIARRLRVTGYDEKRIKQADKLQDIMTMIEQQE TEHA 420
 DANFE I M I EI AGGVRHSEIARRLRVTG+D+ +IKQA+KL+ I+ IE+QE +HA

Sbjct: 361 DANFELITQMPITEINAGGVRHSEIARRLRVTGFD DTKIKQAEKLEQI IETIEKQEAKHA 420

60 Query: 421 YILATYTMLEFREILANHNAIKEM 446
 YILATYTMLEFR +LA+ + + KEM

Sbjct: 421 YILATYTMLEFRSLLADRHVVEKEM 446

-1522-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1388

A DNA sequence (GBSx1473) was identified in *S. agalactiae* <SEQ ID 4255> which encodes the amino acid sequence <SEQ ID 4256>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.3010(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC84011 GB:AF080002 cobyric acid synthase CobQ [Helicobacillus
 mobilis]
 Identities = 89/250 (35%), Positives = 129/250 (51%), Gaps = 9/250 (3%)
20 Query: 11 TKDYKYTLNVAHLYGNLLNTYGDNGNILMMKYVGEKLGCCQMTFDIVSLED RFDPNYYQMA 70
 +K TL + HLY +LLN YGD GNI+ ++ E G + SL ++ + +
 Sbjct: 2 SKTSNRTLTLIHLYPDLLNLYGDRGNIITLRRRCWEGRITLQVHSASLGEKAAFDADLV 61
 Query: 71 FFGGGQDYEQAIVARDLPSKKEDINKFIQNGV-VLAICGGFQLLGQYYIQANGERIEGI 129
 F GGG D EQ ++ +D K G+ +L++CGG+QLLG YY GE + G+
25 Sbjct: 62 FMGGGSDREQTLFQDFQQHKGPALEAAEGGLPILLSVCGGYQLLGLYYRTHTGEEEMPGL 121
 Query: 130 GVMGHYTLNQNNRYIGDIKIHNDEFNE--TYYGFEHQGRFTFLSEDE--KPLGTVIYGN 185
 G+ +T + R IG++ E T GFENH GRFTFL +PL V G
30 Sbjct: 122 GLFDAWT-EAGSTRLIGNVVAQAPLLGEQATLVGFENHSGRTFLGSRGGIQLAQVTAGF 180
 Query: 186 GNNKEDGTEGVHYKNVFGSYFHGPILSRNANLAYRLVATALRNKYG---KEIVLPSYEEI 242
 GNN +D EG YKN G+Y HGP+L +N LA L++ AL +YG + ++E
 Sbjct: 181 GNNGGDQGEQAVYKNAVGTLYLHGPVLPKNPALADWLLSKALERRYGGGSLSTLQDTWENR 240
35 Query: 243 LSLEIPEEYQ 252
 L + + +G
 Sbjct: 241 AHLVAQRFG 250

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4257> which encodes the amino acid
40 sequence <SEQ ID 4258>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45 bacterial cytoplasm --- Certainty=0.2586(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 197/260 (75%), Positives = 224/260 (85%)
 Query: 1 MTYTSLSKSPPTTKDYKYTLNVAHLYGNLLNTYGDNGNILMMKYVGEKLGCCQMTFDIVSLED 60
 MTYTSLSKSP +DY Y L +AHLYGNI+NTYGDNGNILM+KYV EKLG ++T DIVS+ D
 Sbjct: 1 MTYTSLSKSPENQDYIYDLTIAHLYGNLMNTYGDNGNILMLKYVAEKLGARVTVDIVSIND 60
55 Query: 61 RFDPNYYQMAFFGGGQDYEQAIVARDLPSKKEDINKFIQNGVVLAIICGGFQLLGQYYIQ 120
 F+ + Y + FFGGGQDYEQ+IVA+DLPSKK + +I NN VVLAICGGFQLLGQYY+Q
 Sbjct: 61 TFEQDDYDIVFFGGGQDYEQSIVAKDLPSKKAALADYIANNKVVLAIICGGFQLLGQYYVQ 120

-1523-

Query: 121 ANGERIEGIGVMGHYTLNQNNNRYIGDIKIHNDEFNETYYGFENHQGRITFLSEDEKPLGT 180
 ANG +I+G+G+MGHYTLNQ+ NR+IGDIKIHNDEFNETYYGFENHQGRITFLS DEKPLG
 Sbjct: 121 ANGVIDGLGIMGHYTLNQHQNRFIGDIKIHNDEFNETYYGFENHQGRITFLSGDEKPLGR 180

5 Query: 181 VIYGNNGNKKEDGTEGVHYKNVFGSYFHGPILSRNANLAYRLVATALRNKYGKEIVLPSYE 240
 V+YGNNGNKKED TEGVHYKNV+GSYFHGPILSRN NLAYRLV TAL+ KYG I LPSY+
 Sbjct: 181 VVYGNNGNKKEDQTEGVHYKNVYGSYFHGPILSRNVNLAYRLVTTALKKKYGSALSLSYD 240

10 Query: 241 EILSLEIPPEYGDVKSADF 260
 +IL EI EEE D+KSKA F
 Sbjct: 241 DILKQETITEYADLKSKASF 260

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 1389

A DNA sequence (GBSx1474) was identified in *Sagalactiae* <SEQ ID 4259> which encodes the amino acid sequence <SEQ ID 4260>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1701(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04402 GB:AP001509 lipoate-protein ligase [Bacillus halodurans]
 Identities = 153/316 (48%), Positives = 212/316 (66%), Gaps = 3/316 (0%)

30 Query: 10 DPAYNVALEAYAFQKLTIDIDEIFIL-WINEPATIIGRHQNTIQEINKEFIDKNGIHVVRR 68
 DP N+A+E YA + L DI+E ++L +INEP+IIIGR+QNTI+EIN E+++ NGIHVVRR
 Sbjct: 11 DPRINLAIEEYALKNL-DINETYLLFYINEPSIIIGRNQNTIEEINTEYVESNGIHVVRR 69

35 Query: 69 LSGGGAVYHDLNLLNYTIIISNNTQEGAFDFQTFSPKPVIDTLAKLGKAEFTGRNDL-EIN 127
 LSGGGAVYHD NLN++ I+ + E +FQ F+ PVI LAKLGV AE GRND+ +
 Sbjct: 70 LSGGGAVYHDHGNLNFSEFITKDDGESFSNFQKFTDPVIKALAKLGVTAELKGRNDIIASD 129

40 Query: 128 GQKFAGNAQAYYKGRMMHHCILLFDVMSVLGQALKVSKDKIESKGIKSVRARVTNIVDH 187
 G+K +GNAQ KGRM HG LLFD ++ + AL VSKDKIESKGIKS+R+RV NI +
 Sbjct: 130 GRKISGNAQFSTKGRMFSEHGTLLFDSEIDHVVSAIIVSKDKIESKGIKSIRSRVANISEF 189

45 Query: 188 LSDKITVQEFSDAILAQMKEEYEMDEYVLSDAELSEIQAMRDNQFATWDWTYGKAPEYT 247
 L++KI++ +F +L + + + EY L+ + +EI + ++ WDW YGK+P +
 Sbjct: 190 LTEKISIDQFRSLLESIFDQANIQEYKLTADDWAEIHELKSKERYQNWWDWNYGKSPAFN 249

50 Query: 248 IERGVRYPAKITYANVENSTIKSVKIFGDFGVPVDDIEKMLEGVRYDYKDVLAALK 307
 ++ R+P G I V+ TI+ KIFGDFG V D+E L G+RY+ D+ AL
 Sbjct: 250 LQHSRFPVGNIDIRLEVKGGTIQCKIFGDFGCTGDVRDLVDRLVGIRYERADIEQALA 309

Query: 308 TVDTSQYFSRMTPEEI 323
 VD YF ++ ++I
 Sbjct: 310 DVDVKTYFGQVEKDDI 325

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4261> which encodes the amino acid sequence <SEQ ID 4262>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1271(Affirmative) < succ>

-1524-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 249/328 (75%), Positives = 292/328 (88%)

Query: 1 MKYIVNTSNDPAYNVALEAYAFQKLTIDEIFILWINEPAIIIGRHQNTIQEINKEFIDK 60
 MKYIVN S++PA+N+ALEAYAF++L + DE+FILWINEPAIIIG+HQNTIQEINKE+ID+
 10 Sbjct: 1 MKYIVNKSHNPAPFNIALEAYAFRELVEDELFIWINEPAIIIGKHQNTIQEINKEYIDE 60

Query: 61 NGIHVVRRLSGGGAVYHDLNNLNYTIIISNNTQEGAFDFQTFQSKPVIDTLAKLGKAEFTG 120
 +GIHVVRRLSGGGAVYHDLNNLNYTIIISN T EGAFDF+TFS+PVI TLA LGV A FTG
 15 Sbjct: 61 HGIHVVRRLSGGGAVYHDLNNLNYTIIISNKTAEAFDFKTFQSPVIATLADLGVTANFTG 120

Query: 121 RNDLEINGQKFAGNAQAYYKGRMMHHGCLLFVDVMSVLGQALKVSKDKIESKGKSVRR 180
 RND+EI+G+K GNAQAYYKGRMMHHGCLLFVDM+VLG ALKVSVDKIESKG+KSVRR
 20 Sbjct: 121 RNDIEIDGKKICGNAQAYYKGRMMHHGCLLFVDMTVLGDALKVSKDKIESKGKSVRR 180

Query: 181 VTNIIVDHLSDKITVQEFSDAILAQMKBEYPEMDEYVLSDAELSEIQAMRDNQFATWDWTY 240
 VTNI++ L +KITV+EFSD ILA+MKE YP+M EYVLS+ EL++I+ QF +WDWTY
 25 Sbjct: 181 VTNIILNELPEKITVEEFSDKILAKMKETYPDMTEYVLSDELAKIEQSAKEQFGSVDWTY 240

Query: 241 GKAPYTTIERGVRYPAKITYANVENSTIKSVKIFGDFGKVPVDDIEKMLEGVRYDYK 300
 GKAPYTTIER VRYPAKTI+T+ANVENS IK++KI+GDFFG+K V DIE +L G +Y+Y+
 30 Sbjct: 241 GKAPYTTIERNVRYPAKISTFANVENSIIKNLKIYGDFFGIKVDIENLLIGCKYEYR 300

Query: 301 DVLAALKTVDTSQYFSRMTPEEITKAIV 328
 DV LKT+DT+QYFSRMT EE+ KAIV
 35 Sbjct: 301 DVFERLKTIDTTQYFSRMTVEEVAKAIV 328

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1390

A DNA sequence (GBSx1475) was identified in *S.agalactiae* <SEQ ID 4263> which encodes the amino acid sequence <SEQ ID 4264>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.70 Transmembrane 294 - 310 (294 - 312)

40 ----- Final Results -----

bacterial membrane --- Certainty=0.1680 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA21748 GB:L31844 dihydrolipoamide dehydrogenase [Clostridium magnum]
 Identities = 229/589 (38%), Positives = 339/589 (56%), Gaps = 25/589 (4%)

50 Query: 1 MAFDVIMPKLGVDMEGEILEWKKNEGDTVNEGDLLEIMSDKTNMEIEAEDTGVLKIV 60
 MA V+MPKLG+ M EG ++ WKK EGD V G++L E+ +DK E+E+ D G++ K++
 Sbjct: 1 MAKIVVMPKLGLTMTGETLVTWKKAEQDQVKVGEILFEVSTDKLTNEVESSEDEGIVRKL 60

Query: 61 HQAGDVVPVTEVIAYIGEEGEEVGTSSPSADATITAEDGQSVSGPAAPSQETVAAATPKE 120
 GDVV +A IG E++ + +G S +A +T A PK+
 55 Sbjct: 61 VNEGDDVVECLNPVAIIGSADEDISSLL-----NGSSESGSGSARQSDTKA---PKK 107

Query: 121 ELAADEY--DIVVVGPGPAGYYAAIRGAQLGGKIAIVEKTEFGGTCLNVGCIPTKTYLKN 178
 E+ A + ++VV+GGGP GY AAIR AQLG K+ ++EK GGTCLNVGCIPTK L +
 60 Sbjct: 108 EVEAVKGGDNLVVGPGPGYVAAIRAAQLGAKVTLIEKESLGGTCLNVGCIPTKVLLHS 167

-1525-

Query: 179 AEILDGLKVAAGRGINLASTNYAIDMDKTVAFKNSVVKTLTGGVRGLLKANKVEIFNGLG 238
 +++L +K GI++ + ++ K V+K L GV GLL NKV++ G
 Sbjct: 168 SQLLTEMKEGDKLGIDIEGS-IVVNWKHIQKRKKIVIKLVSGVSGLLTCNKVKVIKGT 226

5 Query: 239 QVNPDKSVVIGDK-----VIKGRNVVLATGSKVSRINIPGIESPLVLTSDDILDREIPK 293
 + ++++ + + N ++ATGS I G + V+ S L L P+
 Sbjct: 227 KFESKDTILVTKEGVAEKVNFDAIATGSMPPFIPIEIGNKLSGVIDSTGALSLESNPE 286

10 Query: 294 SLAVMGGGVVGIELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKILAKKGMKIKT 353
 S+A++GGGV+G+E ++ S G V++IEM I+P MD+E+S + L + G+ I +
 Sbjct: 287 SIAIIGGGVIGVEFASIFNSLGCKVSIIEMLPHILPPMDREISEIAKAKLIRDGININN 346

15 Query: 354 VGVSEIVEANNQLTLKL--NNGEEVV-ADKALLSIGRVPQMNGLENLEPELEMERGRIK 410
 V+ I + + L + + GEE + +K L+++GR + GL+ + ++ E G I V
 Sbjct: 347 CKVTRIEQGEDGLKVSFIGDKGEESIDVEKVLIAVGRRSNIEGLDVEKIGVKTEGGSIIV 406

20 Query: 411 NAYQETSIPGIYAPGDVNGTRMLAHAAYRMGEVAAENALGGNKRKAHLDFTPAAVYTHPE 470
 N ET++ GIYA GD G MLAH A G VAAEN +G NK K PA VYT PE
 Sbjct: 407 NDKMETNVEGIYAIGDCTGKIMLAHVASDQGVVAAENIMGQNK-KMDYKTVPACVYTKPE 465

25 Query: 471 VAMVGMTTEEQAREQYGDILVGNKSFNGRAIASNEAHGFVKVIAEPKYKEILGVHIIGP 530
 +A VG+TEEQA+E+ D VGK NG+++ NE G +K+I + KY+EILGVHI+GP
 Sbjct: 466 LASVGLTEEQAKEKGIDYKVGKFLAANGKSLIMNETGGVIKIITDKKYEEILGVHILGP 525

Query: 531 AAAELINEASTIMENELTVYDVAQSIHGHPHFSEVMYEAFLDVLGEAII 579
 A +LI EA+ + E T+ ++ ++H HPT E M EA L V +AIH
 Sbjct: 526 RATDLITEAALALRLAATLEEIITTVHAHPTVGEAMKEAALAVNNQAIH 574

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1819> which encodes the amino acid
 30 sequence <SEQ ID 1820>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.70 Transmembrane 297 - 313 (297 - 315)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 497/591 (84%), Positives = 538/591 (90%), Gaps = 10/591 (1%)

Query: 1 MAFDVI MPKLGVD MQE G E I L E W K K N E G D T V N E G D V L L E I M S D K T N M E T E A E D T C V L L K I V 60
 MA ++IMPKLGVD MQE G E I + E W K K E G D T V N E G D + L L E I M S D K T N M E + E A E D + C V L L K I .
 45 Sbjct: 1 MAVEI I MPKLGVD MQE G E I L E W K K Q E G D T V N E G D I L L E I M S D K T N M E L E A E D S G V L L K I T 60

Query: 61 HQAGDVVPVTEVIAYIGEEGEEVGTSSPSA---DATITAEDGQS--VSGPAAPSQETVAA 115
 QAG+ VPVTEVI YIG EGE V SSP+A + T ED ++ + P AP+Q A+
 50 Sbjct: 61 RQAGETVPVTEVIGYIAGGESVEVSSPAASDVNVARTTEDLEAGLEVPKAPAQ--AAS 118

Query: 116 ATPKEELAADDEYDIVVGGGPAGYAAIRGAQLGGKIAIVEKTEFGGTCLNVGCIPTKTY 175
 A PK LA DEYDI+VVG GGPAGYAAIRGAQLGGKIAIVEK+EFGGTCLNVGCIPTKTY
 Sbjct: 119 AAPKAALADDEYDII VVG GGPAGYAAIRGAQLGGKIAIVEKSEFGGTCLNVGCIPTKTY 178

55 Query: 176 LKNAEILDGLKVAAGRGINLASTNYAIDMDKTVAFKNSVVKTLTGGVRGLLKANKVEIFN 235
 LKNAEILDG+K+AAGRGINLASTNY IDMDKT V FKN+VVKTLTGGV+GLLKANKV IFN
 Sbjct: 179 LKNAEILDGIKIAAGRGINLASTNYTIDMDKTVDKNTVVKTLTGGVQGLLKANKVTIFN 238

60 Query: 236 GLGQVNPDKSVVIGDKVIKGRNVVLATGSKVSRINIPGIESPLVLTSDDILDREIPKSL 295
 GLGQVNPDK+V IG + IKGRNV+LATGSKVSRINIPGI+S LVLTSDDILDRE+PKSL
 Sbjct: 239 GLGQVNPDKTITIGSQTIKGRNVILATGSKVSRINIPGIDSKLVLTSDDILDREMPKSL 298

Query: 296 AVMGGGVVGIELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKILAKKGMKIKTSVG 355
 AVMGGGVVGIELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKIL+KKGMIKTSVG
 65 Sbjct: 299 AVMGGGVVGIELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKILSKKGMKIKTSVG 358

-1526-

5 Query: 356 VSEIVEANNQLTLKLNNGEEVVADKALLSIGRVPQMNGLENLEPELEMERGRIKVNAYQE 415
 VSEIVEANNQLTLKLNNGEEVVA+KALLSIGRV QMNGLENL LEM+R RIKVN YQE
 Sbjct: 359 VSEIVEANNQLTLKLNNGEEVVAEKALLSIGRVSQMNGLENL--NLEMDRNRIKVNNDYQE 416

10 Query: 416 TSIPGIYAPGDVNGTRMLAHAAYRMGEVAAENALGGN-KRKAHLDFTPAAVYTHPEVAMV 474
 TSIPGIYAPGDVNGT+MLAHAAYRMGEVAAENA+ GN RKA+L +TPAAVYTHPEVAMV
 Sbjct: 417 TSIPGIYAPGDVNGTKMLAHAAYRMGEVAAENAMHGNTTRKANLKYTPAAVYTHPEVAMV 476

15 Query: 475 GMTTEEQAREQYGDILVGKNSFTGNGRAIASNEAHGFKVIAEPKYKEILGVHIIGPAAAE 534
 G+TEEQAREQYGD+L+GKNSFTGNGRAIASNEAHGFKVIA+ KY EILGVHIIGPAAAE
 Sbjct: 477 GLTEEQAREQYGDVLIGKNSFTGNGRAIASNEAHGFKVIADAKYHEILGVHIIGPAAAE 536

Query: 535 LINEASTIMENELTVYDVAQSIHGHPFSEVMYEAFLDVLGEAIHNPPKRK 585
 +INEA+TIME+ELTV ++ SIHGHPFSEVMYEAFLDVLGEAIHNPPKRK
 Sbjct: 537 MINEAATIMESELTVDLELLLSIHGHPFSEVMYEAFLDVLGEAIHNPPKRK 587

SEQ ID 4264 (GBS681) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 165 (lane 2; MW 68.3kDa) and in Figure 188 (lane 10; MW 68kDa).

20 Purified GBS681-His is shown in Figure 240, lane 5-6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1391

25 A DNA sequence (GBSx1476) was identified in *S.agalactiae* <SEQ ID 4265> which encodes the amino acid sequence <SEQ ID 4266>. This protein is predicted to be dihydrolipoamide acetyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4466(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04497 GB:AP001509 dihydrolipoamide S-acetyltransferase
 [Bacillus halodurans]
 Identities = 187/462 (40%), Positives = 266/462 (57%), Gaps = 26/462 (5%)

40 Query: 1 MAVELIMPKLGVDMQEGEILEWKKQVGDVNEGDLLEIMSDKTNMEIEARDSGVLLKIT 60
 MA EI MPKL MQEG +L+W K+ GD V G+ L EIM+DK N+E+EA + G LLK
 Sbjct: 1 MAKEIFMPKLSSTMQEGTLLQWFKKEGDRVEVGEPLFEIMTDKINIEVEAYEEGTLLKRY 60

45 Query: 61 HGNGDVVPVTETIGYIGAEGREVTEASSSENTSVEENATQVTSEPEKVEETSEPSVPAAT 120
 +G D +PV IGYIG E V +E E T E T+ P++
 Sbjct: 61 YGEDDEIPNVHVIYIGTIPDESVP----TEPPGASEITASSTDEAGDHRTTAVKKAPSSD 116

50 Query: 121 SGEKVRATPAARKLAREMSIDLALVSGTGANGRVHREDVENFKGAQPRITPLARRIAEDQ 180
 E VRATPAAR++A+E IDL V G+G GRV DV FK + TPLA+++AE +
 Sbjct: 117 R-ENVRATPAARRIAKEKRIDLRLQVEGSGPEGRVQAVDVATFKKKGQKATPLAKKVAEVK 175

55 Query: 181 GVDIAEITGSGIRGKIVKNDVLAAMSPQAAEAPVETKATPTTTEKQLPEGVEVIKMSAMR 240
 GV + ++ GSG GK+ + DV A A +PVE K +K+S +R
 Sbjct: 176 GVALEKVGSGPYGKVYREDVEHAQ-----AASPVEDKGNR-----VKLSGLR 218

Query: 241 KAISKGMTNSYLTAPSFTLNYDIDMTEMMALRKKLIDPIMAKTGLKVSFTDLIGMAVVKI 300
 K ++K M +S +AP T+ +IDM+ + +R +L+ I +TG ++S+T+++ AV
 Sbjct: 219 KVVAKRMVDSAFSAPHVTITTEIDMSSTIKIRSQLLGMIEQETGYRLSYTEIVMKAVAHA 278

-1527-

Query: 301 LMKPEHRYLNASLINDAQEIELHNFVNIGIAGVGLDDGLIVPVVHNADQMSLSDFVIASKD 360
 LM H +NAS + EI H V+IG+AV ++ GL+VPVV + D+ L+ K
 Sb|ct: 279 LMS--HPTINASFFEN--EIVYHEDVHIGLAVAVEGGLVVPVVKHVDDKGLAQLTNECKT 334

Query: 361 VIKKTQEGKLSAEMSGSTFSITNLGMFGTKTFNPIINQPN SAILGVGATIPTPTVVDGE 420
 V ++ +L MSG TF+I+NLGM+ F P+INQP SAILGVG P +DG+
 Sb|ct: 335 VAMAARDNRLSQEMMSGTFTISNLGMYAIDVFTPVINQPN SAILGVGRIQEKPVGIDGQ 394

Query: 421 IVARPIMAMCLTIDHRIVDGMNGAKFMVDLKNLMENPFGLLI 462
 I RP+M L+ DHR++DG A F+ D+K+++E PF LL+
 Sb|ct: 395 IELRPMMTASLSFDHRVIDGAPAAFLTDVKSMLQPFQLIM 436

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4267> which encodes the amino acid sequence <SEQ ID 4268>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4774(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 354/473 (74%), Positives = 390/473 (81%), Gaps = 15/473 (3%)

Query: 1 MAVELIIMPKLGVD MQEGEILEWKKQVGDV VNEG DVLLBIMSDKTNMEIEAEDSGVLLKIT 60
 MA EIIMPKLGVD MQEGEI+EWKKQ GD VNEG D+LLEIMSDKTNME+EAEDSGVLLKIT
 Sb|ct: 1 MAFELIIMPKLGVD MQEGEIIEWKKQEGD TVNEG DILLEIMSDKTNMELEAEDSGVLLKIT 60

Query: 61 HGNGDVVPVTE TIGYIGAEGEEVTEASSSENTS-----VEENATQVTSEPEKVEETSEPS 115
 GD VPVTE IGYIGAEGE V +SSE T+ +A + E V + P
 Sb|ct: 61 RQAGD TVPVTEVIGYIGAEGESVD TIASSEKTEIPVFPASADAGPAVAPKENVASPA-PQ 119

Query: 116 VPAAT----SGEKVRATPAARKLAREMSIDLALVSGTGANGRVHREDVENFKGAQPRITP 171
 V A +G KVRATPAARK A EM IDL V CTG GRVH+EDVENFKGAQP+ +P
 Sb|ct: 120 VAATAIPQNGGKVRATPAARKAAEMGIDL GQVPGTGPKGRVHKEDVENFKGAQPKASP 179

Query: 172 LARRIAEDQGVDAIEITGSGIRGKIVKNDVLAAMSPQAAEAPVETKATPTTEEK--QLPE 229
 LAR+IA D+G+D+A ++G+G GK++K D++A + A P E KA EEK LPE
 Sb|ct: 180 LARKIAADK GIDLATVSGTGFGNKVMKEDIMAILL---AAKPAEAKAPAAKEEKVVDLPE 236

Query: 230 GVEVIKMSAMRKAISKGMTNSYL TAPSF TNLNYDIDMTEMALRKKLIDPIMAKTGLKV SF 289
 GVE MSAMRKAISKGMTNSYL TAP+FTNLNYDIDMTEM+ALRKKLIDPIMAKTGLKV SF
 Sb|ct: 237 GVEHKPMSAMRKAISKGMTNSYL TAPTFTNLNYDIDMTEMIALRKKLIDPIMAKTGLKV SF 296

Query: 290 TDLIGMAVVKTLMKPEHRYLNASLINDAQEIELHNFVNIGIAGVGLDDGLIVPVVHNADQM 349
 TDLIGMAVVKTLMKPEH Y+NASLINDA +IELH FVN+GIAGVGLDDGLIVPV+H A++M
 Sb|ct: 297 TDLIGMAVVKTLMKPEHEYMNASLINDANDIELHRFVN LGIAGVGLDDGLIVPVIHGANKM 356

Query: 350 SLSDFVIASKDVIKKTQEGKLSAEMSGSTFSITNLGMFGTKTFNPIINQPN SAILGVGA 409
 LSDFV+ASKDVIK Q GK LK+AEMSGSTFSITNLGMFGTKTFNPIINQPN SAILGVGA
 Sb|ct: 357 CLSDFVLASKDVIKKAQTGKLKAAEMSGSTFSITNLGMFGTKTFNPIINQPN SAILGVGA 416

Query: 410 TIPTPTVVDGEIVARPIMAMCLTIDHRIVDGMNGAKFMVDLKNLMENPFGLLI 462
 TIPTPTVVDGEIV+RPIMAMCLTIDHR+VDGMNGAKFMVDLKNLMENPF LLI
 Sb|ct: 417 TIPTPTVVDGEIVSRPIMAMCLTIDHRLVDGMNGAKFMVDLKNLMENPFELLI 469

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1528-

Example 1392

A DNA sequence (GBSx1477) was identified in *S.agalactiae* <SEQ ID 4269> which encodes the amino acid sequence <SEQ ID 4270>. This protein is predicted to be acetoin dehydrogenase (TPP-dependent) beta chain (pdhB). Analysis of this protein sequence reveals the following:

5 Possible site: 18
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.1267(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9779> which encodes amino acid sequence <SEQ ID 9780> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

 >GP:BAB04496 GB:AP001509 acetoin dehydrogenase (TPP-dependent) beta
 chain [Bacillus halodurans]
 Identities = 189/319 (59%), Positives = 249/319 (77%), Gaps = 1/319 (0%)

20 Query: 11 EAINVAMSEEMRKDEKVFILMGEDVGVGDFGTSVGMLEEFCAKRVDRTPISEAAIAGSA 70
 EAI AM+ EMRK+E VF++GED+GVYGG FG + GM+EEFG++RVR+TPISEAAI+G+A
 Sbjct: 8 EAIREAMTLEMRKNEDVFILGEDIGVYGAFVIRGMIEEFGSEVRNTPISEAAISGTA 67

25 Query: 71 IGAAQTGLRPIVDLTFMDFVTIAMDIAIVNQAKTNYMFGGGLSTPVTFRVASGSGIGSAA 130
 IGAA TG+RPI++L F DF+TIAMD +VNQ AK YM+GG P+ R +GSG G+AA
 Sbjct: 68 IGAAITGMRPILELQSFITIAMDNMVNQAAKLRYMYGGKAKVPMVLRTPAGSGTGAAA 127

30 Query: 131 QHSQSLEAWLTHIPGLKVVAPGTVNESKALLKSSILDNNPVIFLEPKALYGKKEEVNMDP 190
 QHSQSLEAW+THIPGLKVV P T ++K LLK++I DNNPVIF E K Y K V +
 Sbjct: 128 QHSQSLEAWMTHIPGLKVQPATAYDAKGLLKAAIDNNPVIFYEHKLCYRTKCHV-PEE 186

35 Query: 191 DFIYPLGKGDIKREGTDLTIVSYGRMLERVMQAAEEVAAEEGINVEVVDPRTLIPLDKELI 250
 ++ IPLGK D+KR+GTD+T+V+ M+ + ++AA E+ +EGI+VEV+DPRTL+PLD+E I
 Sbjct: 187 EYSIPLGKADVKKRGTDVTVVATAVMVHKALEAAVELEKEGISVEVIDPRTLVPLDEETI 246

40 Query: 251 IDSVKKTGKLILVNDAYKTGGFTGEIATMVAESEAFDYLDHPHIVRLASEDVPVPYSRVLE 310
 I SVKKT +LI+V++A K GGF GEIA+++AESEAFDYLD PI RL + VP+PY+ LE
 Sbjct: 247 IRSVKKTSRLIVVHEAVKRGFGGEBIASIAESEAFDYLDAPIKRLGGKPVPIPNPTLE 306

45 Query: 311 QGILPDVAKIKDAIYKVVN 329
 + +P V I +A+ + +N
 Sbjct: 307 RAAIPQVPDIIIEAVKETLN 325

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4271> which encodes the amino acid sequence <SEQ ID 4272>. Analysis of this protein sequence reveals the following:

 Possible site: 18
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.00 Transmembrane 81 - 97 (81 - 97)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the databases:

 >GP:BAB04496 GB:AP001509 acetoin dehydrogenase (TPP-dependent) beta
 chain [Bacillus halodurans]
 Identities = 187/319 (58%), Positives = 244/319 (75%), Gaps = 1/319 (0%)

-1529-

Query: 11 EAVNLAMTEEMRKDENIFLMGEDVGVYGGDFGTSVGMIEEFGPKRVKDTPISEAAISGAA 70
EA+ AMT EMRK+E++F++GED+GVYGG FG + GMIEEFG +RV++TPISEAAISG A
Sbjct: 8 BAIREAMTLEMARNEDVFIILGEDIGVYGGAFGVTRGMIEEFGSERVRNTPISEAAISGTA 67

5 Query: 71 IGAAITGLRPIDVTFMDFLTIMMDAIVNNGAKNNYMFGGGLITPVTFRVASGSGIGSAA 130
IGAA+TG+RPI+++ F DF+TI MD +VN AK YM+GG P+ R +GSG G+AA
Sbjct: 68 IGAAITGMRIPIELQFSDFITIAMDNMNVQAALRYMYGGKAKVPMVLRTPAGSGTGAAA 127

10 Query: 131 QHSQSLEAWLTHIPGIKVVAPGNANDAKGLLKSARDNNIVLFMEPKALYGGKEEVNODP 190
QHSQSLEAW+THIPG+KVV P A DAKGLLK+AI DNN V+F E K Y K V ++
Sbjct: 128 QHSQSLEAWMTHIPGLKVVQPATAYDAKGLLKAIDNNPVIFYEHKLCYRTKCHVPPE- 186

15 Query: 191 DFYIPLGKGDIKREGTDLTIVSYGRMLERVLQAAEEVAADGINVEVDPRTLIPLDKELI 250
++ IPLGK D+KR+GTD+T+V+ M+ + L+AA E+ +GI+VEV+DPRTL+PLD+E I
Sbjct: 187 EYSIPLGKADVVRKGTDTVTAVATAMVHKALEAAVELEKEGISVEVIDPRTLVLDEETI 246

20 Query: 251 IESVKKTKGLMLVNDAYKTGGFGEIATMITSEAFDYLDHPVRLASEDVPVPYARVLE 310
I SVKKT +L++V++A K GGF GEIA++I ESEAFDYLD PI RL + VP+PY LE
Sbjct: 247 IRSVKKTSRLIVVHEAVKRGFGGEIASIIAEESEAFDYLDAPIKRLGGKPVPIPNPTLE 306

Query: 311 QAILPDVEKIKAAIVKMAN 329
+A +P V I A+ + N
Sbjct: 307 RAAIPQVPDIEAVKETLN 325

25 An alignment of the GAS and GBS proteins is shown below.
Identities = 286/331 (86%); Positives = 310/331 (93%)

Query: 1 MSETKVMALREAINVAMSEEMRKDEKVFMLGEDVGVYGGDFGTSVGMLEEFGAKRVDRTP 60
MSETK+MALREA+N+AM+EEMRKDE +FLMGEDVGVYGGDFGTSVGM+REFG KRV+DTP
30 Sbjct: 1 MSETKLMALREAVNLAMTEEMRKDENIFLMGEDVGVYGGDFGTSVGMIEEFGPKRVKDTP 60

Query: 61 ISEAAIAGSAIGAQTGLRPIDVTFMDFVTIAMDAIVNQAKTNYMFGGGLSTPVTFRV 120
ISEAAI+G+AIGAA TGLRPIDV+TFMDF+TI MDAIVN GAK NYMFGGGL TPVTFRV
35 Sbjct: 61 ISEAAISGAAIGAATGLRPIDVTFMDFLTIMMDAIVNNGAKNNYMFGGGLITPVTFRV 120

Query: 121 ASGSGIGSAAQHSQSLEAWLTHIPGLKVVAPGTVNESKALLKSSILDNNPVIFLEPKALY 180
ASGSGIGSAAQHSQSLEAWLTHIPG+KVVAPG N++K LLKS+I DNN V+F+EPKALY
40 Sbjct: 121 ASGSGIGSAAQHSQSLEAWLTHIPGIKVVAPGNANDAKGLLKSARDNNIVLFMEPKALY 180

Query: 181 GKKEEVNMDPDFYIPLGKGDIKREGTDLTIVSYGRMLERVMQAAEEVAEEGINVEVDPR 240
GKKEEVN DPDFYIPLGKGDIKREGTDLTIVSYGRMLERV+QAAEEVA +GINVEVDPR
45 Sbjct: 181 GKKEEVNMDPDFYIPLGKGDIKREGTDLTIVSYGRMLERVLQAAEEVAADGINVEVDPR 240

Query: 241 TLIPLDKELIIDSVKKTGKLIIVNDAYKTGGFTGEIATMVAESEAFDYLDHPVRLASED 300
TLIPLDKELII+SVKKTGKL+LVNDAYKTGGF GEIATM+ ESEAFDYLDHPVRLASED
50 Sbjct: 241 TLIPLDKELIIESVKKTGKLIIVNDAYKTGGFGEIATMITSEAFDYLDHPVRLASED 300

Query: 301 VPVPYSRVLEQGILPDVAKIKDAIKVNVNKG 331
VPVPY+RVLEQ ILPDV KIK AI K+ NKG
Sbjct: 301 VPVPYARVLEQAILPDVEKIKAAIVKMANKG 331

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1393

55 A DNA sequence (GBSx1478) was identified in *Sagalactiae* <SEQ ID 4273> which encodes the amino acid sequence <SEQ ID 4274>. Analysis of this protein sequence reveals the following:

Possible site: 48
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -3.03 Transmembrane 161 - 177 (161 - 178)
60 ----- Final Results -----
bacterial membrane --- Certainty=0.2211(Affirmative) < succ>

-1530-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9777> which encodes amino acid sequence <SEQ ID 9778> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP: BAB04495 GB: AP001509 acetoin dehydrogenase (TPP-dependent) alpha chain [Bacillus halodurans]
 Identities = 148/317 (46%), Positives = 214/317 (66%), Gaps = 1/317 (0%)

Query: 8 LSKEQHLDMLKMQIRVDVDMKFNKLVRRGFVQGMTHFSVGEEAASVGAIQDLTDSDIIF 67
 +++++ +D+F +M IR + K ++ +G + G TH +VG+EA++VG+I L + D +
 Sbjct: 10 MTEKKLVDLFKQMWLIRYFEEKVDEFFAKGMIHGTTHLAVGQEASAVGSIAVLEERDKLT 69

Query: 68 SNHRGHGQTIAGKIDIGGMFAELAGKATGTSKGRGGSMLANLEKGNYGNGIVGGGYAL 127
 S HRGHG IAKG D+ M AEL G+ TG KG+GGSMH+A++E+GN G NGIVGGG+++
 Sbjct: 70 STHRGHGHCIAKGADVNRMMAELEFRETGYCKGKGGSMHIADVERGNLGANGIVGGGFSI 129

Query: 128 AVGAALTQQYEGTDNIVIAFSGDSATNEGSFHESVNLAAVWNLPIVFFIINNRYGISTDI 187
 A GAALT + + +V+ F GD A+NEGSFHE+VNLA++W LPV+F NN+YG+S +
 Sbjct: 130 ATGAALTSMKKEGYVLCFFGDGASNEGSFHEAVNLASIWKLPPVFICENNQYMSGGSV 189

Query: 188 TYSTKIPHYLMRADAYGIPGHYVEDGNDLMAVYEKMHEVINYVRSGNGPAIVEVESYRWF 247
 I H+ RA YGIPG V DGND+ AV + ++ R G GP IVE ++YRW
 Sbjct: 190 KEMINIEHISDRAAGYIGP-GMVVDGNDVFAVMNVVGRAVDRARRGEGPTIVEAKTYRWK 248

Query: 248 GHSTADAGVYRTKEEVDSWKAKDPVKRYRAYLIENEIATEEELAAIEAQVIKEVEEGVKF 307
 GHS +DA YRT+EE W+ KDP+ R RA L++ I TEEE +I+ + +++E+ V+F
 Sbjct: 249 GHSKSDAKKYRTREEKEWREKDPARLRATLVKEGIVTEEEADSIQBEAKQKIEDSVQF 308

Query: 308 AESPPFDMSVAFEDVF 324
 A SP P++ EDV+
 Sbjct: 309 ARNSPEPEIESLLEDVY 325

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4275> which encodes the amino acid sequence <SEQ ID 4276>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3502 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 244/326 (74%), Positives = 278/326 (84%)

Query: 1 MEVRMVTLSKEQHLDMLKMQIRVDVDMKFNKLVRRGFVQGMTHFSVGEEAASVGAIQDL 60
 ME MVT+SKEQHLDMLKM+RIR+ D + NKLVRRGFVQGMTHFSVGEEAA+VGA+ L
 Sbjct: 1 MEAEMVTVSKEQHLDMLKMERIRBFDNRINKLVRRGFVQGMTHFSVGEEAANVGAVAH 60

Query: 61 TDSDIIFSNHRGHGQTIAGKIDIGGMFAELAGKATGTSKGRGGSMLANLEKGNYGNGI 120
 + DIIFSNHRGHGQ+IAK +D+ M AELAGKATG SKGRGGSMLA+ EKGNYGTNGI
 Sbjct: 61 SYDDIIFSNHRGHGQSIKMDMLNKMAELAGKATGVSKGRGGSMLADFEKGNYGNGI 120

Query: 121 VGGGYALAVGAALTQQYEGTDNIVIAFSGDSATNEGSFHESVNLAAVWNLPIVFFIINN 180
 VGGGYALAVGAALTQQY+GT+NI +AFSGD ATNEGSFHESVN+AA W LPVFFIINN
 Sbjct: 121 VGGGYALAVGAALTQQYKGTNNIAVAFSGDGATNEGSFHESVNMAATWKLPIVFFIINN 180

Query: 181 YGISTDITYSTKIPHYLMRADAYGIPGHYVEDGNDLMAVYEKMHEVINYVRSGNGPAIVE 240
 YGIS I +T PHLY RA+AYG+PG Y EDGND+MAVYE M + + +VR GNGPAIVE
 Sbjct: 181 YGISMSINNATNTPHLYTRAEAYGVPGFYCEDGNDVMAVYETMGKAVEHVRGGNGPAIVE 240

```

Query: 241 VESYRWFGHSTADAGVYRTKEEVDSWKAKDPVKRYRAYLIENEIATEBELAAIEAQVIKE 300
          VESYRWFGHSTADAG YRTKEEVD WK KDP+ +YR YL IAT++EL AI+AQV KE
Sbjct: 241 VBSYRWFGHSTADAGKYRTKEEVDWEKDKPMIKYRTYLTSEGIATDDELDAIQAVQVKE 300

Query: 301 VEEGVKFAEESPFPDMSVAFEDVFDV 326
          V++ +FA+ SP P++SVAFEDV+VD
Sbjct: 301 VDDAYEFAONSPDPELSVAFEDVWD 326

```

```

Lipop: Possible site: -1    Crend: 10
McG: Discrim Score:      -14.75
GvH: Signal Score (-7.5): -4.24
      Possible site: 48
>>> Seems to have no N-terminal signal sequence
ALOM program    count: 1 value: -3.03 threshold: 0.0
      INTEGRAL    Likelihood = -3.03    Transmembrane 161 - 177 ( 161 - 178)
      PERIPHERAL  Likelihood = 3.55      117
modified ALOM score:      1.11

*** Reasoning Step: 3

```

The protein has homology with the following sequences in the databases:

231 261 291 321 351 381 411 441
F*IEMFFTKTKKAVQILASCEKNLYNN*VIKIFLEVRMVLTSKEQHLDMLFKMQIRIDVDMKFNLKLVRRGFVGQMTHFSV
 :: : : :: : : | : || || : : : : : : : :
 MKLLKREGLSLTEEKALWMIYQKMLEIRGFDKVHELFAQGVLPGFVHLIYA
 10 20 30 40 50

711 741 771 801 831 861 891 921
AVGAALTQQYEGTDNIVIAFGSDSATNEGSFHESVNLAAYVWNLVPVFFIINNRYGISTDITYSTKIPHLYMRADAYGIGP
| :||| :| : | : | : ||| :||| ||| :| ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
ACGSALTAKYQKTKNVSVCFFGDGANNOGTTFHEGLNLAAYVWNLPPVFVAENNGYGGEATPFHEYASACDSIADRAAAYNMFG

140 150 160 170 180 190 200 210

60 951 981 1011 1041 1071 1098 1128 1158
HYVEDGNDLMAVYEKMHEVINIVYRSGNGPAIVEVESYRWFHGHSTADAGVYRTKEE-VDSWKAKDPVKRYRAYLIENEIAT
|| ::||: | | :| ||::: :|| :|| || :||:| :| :| :: :| ||:: |
-VTVDGKDILAVYQAAEEAIERARNRGGGSLIECMTRYNGHYFEGDAQTYTKKDERVEHLEEKDAIQGFKNYLLKETDAN
220 230 240 250 260 270 280

-1532-

```

1188      1218      1248      1278      1308      1338      1368      1398
EEELAAIEAQVIKEVEEGVKFAEESPPFDMSVAFEDVFDV*NNLK*MRFISFYYSID*KTDIRRK**AKLKLWLCAKRLM
:|: || :|: :|: | |:|:|:| | : ||:|
--KLSDIEQRVSESEIKAVSFSEDSYPKDSSELLTDVYVSYEKGGM
300      310      320      330

```

SEQ ID 8798 (GBS403) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 2; MW 64.4kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 4; MW 39.5kDa).

GBS403-GST was purified as shown in Figure 218, lane 6.

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1394

A DNA sequence (GBSx1479) was identified in *S.galactiae* <SEQ ID 4277> which encodes the amino acid sequence <SEQ ID 4278>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2464(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9775> which encodes amino acid sequence <SEQ ID 9776> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB12414 GB:Z99107 similar to ABC transporter (ATP-binding
protein) [Bacillus subtilis]
Identities = 328/643 (51%), Positives = 443/643 (68%), Gaps = 9/643 (1%)

Query: 9 MIILQGNKIERSFSGDVLFDNINIQVDQDRDIALVGRNGAGKSTLLKILVGEEAPTKEGI 68
M+ILQ N++ +SF D + +NI ++V RDRIA+VGRNGAGKSTLLKI+ G+ + KGEI
Sbjct: 1 MMILQANQLSKSFGADTILNNIKLEVRNRDRIAIVGRNGAGKSTLLKIIAGQLSYEKGEI 60

Query: 69 NKKRDLSSYLAQDSRFQSENTIFQEMLQVFDLSLREVEKRLRELELQMGQVSGSDLEQLM 128
K +D+++ YLAQ + S+ TI +E+L VFD L+ +EK +R +E +M +LE +M
Sbjct: 61 IKPKDITMGYLAQHTGLDSKLTIKEBLTIVFDHLKAMEKEMRAMEEKMAAADPGELESIM 120

Query: 129 KTYDILSEEFREKGGFTYESDIKAILNGFKFNSDMWEMPISELGGQNTRLALAKMLLEK 188
KTYD L +EF++KGG+ YE+D++++L+G F+ + LSGGQ TRIAL K+LL +
Sbjct: 121 KTYDRLQQEFKDKGGYQYEAADVRSVLHGLGFHFDSTQVQSLGGQKTRLALGKLLLTQ 180

Query: 189 PELLVLDEPTNHLIDITIAWLENYLVNYQGALIIVSHDRYFLDKVATVTYDLTTHSLDRY 248
P+LL+LDEPTNHLIDIT+ WLE+YL Y GA++IVSHDRYFLDKV Y+++ +Y
Sbjct: 181 PDLILLDEPTNHLIDITLTWLEHYLQGYSGAILIVSHDRYFLDKVVNOVYEVSRAESKCY 240

Query: 249 VGNYSKFMDLKAEKIAATEEKNFEKQKEIAKLEDFVQRNIVRASTTKRAQARRKQLEKME 308
GNYS ++D KA + + K +EKQQ EIAKL+DFV RN+ RASTTKRAQ+RRKQLE+M+
Sbjct: 241 HGNYSAYLDQKAAQYKDLKMYEKQDEIAKLQDFVDRNLARASTTKRAQSRRKQLERMD 300

Query: 309 RLDKPNVEQKSANMTFHAGKVGSGNVVLTLENAAIGYEG-VSLSEPIDLDVKKFDAIAIVG 367
+ KP ++KSAN F K SGN VL +++ I YE L + + + ++ A+VG
Sbjct: 301 VMSKPLGDEKSANFHFDTIKQSGNEVLRVQDLTISYENQPPLLTEVSFMLTRGESAAVLG 360

Query: 368 PNGIGKSTLIKSLVGGQIPFIKGEAKLGANVETGYDQSQSNLTKTNTVLDELWDFAFSTTP 427
PNGIGKSTL+K+L+ + +G G+NV GYYDQ Q+ LT + VLDELWD + P

```

-1533-

Sbjct: 361 PNGIGKSTLLKTLIDTLKPDQGTISYGSNVSVGYDQEQAEITSSKRVLDELWDEYPLGLP 420

Query: 428 EVEIRNRLGAFLESGDDVKKSVSMLSGGERARLLAKLSMENNFLILDEPTNHLDDISK 487
E EIR LG FLFSGDDV K V LSGGE+ARL LAKL ++ NFLILDEPTNHLID+DSK

5 Sbjct: 421 EKEIRTCLGNFLFSGDDVLKPVHSLSGGEKARLALAKLMLQKANFLILDEPTNHLDDISK 480

Query: 488 EVLENALIEFDGTLFVSHDRYFINRVATKVLEISDKGSTLYLGDYDYLYTKKAELEELA 547
EVLENALI++ GTLLFVSHDRYFINR+AT+VLE+S YLGDYDYY KK E EL

10 Sbjct: 481 EVLENALIDYPGTLFVSHDRYFINRIATRVLELSSSHIEEYLG DYDYYTEKKTEQLELE 540

Query: 548 RLNEEEVSASKTEIDVTS-----YETQKANQKEFRKITRRVVEIEARLEVLNENNING 603
++N++E KT V SD YE +K +K+ R+ RR+ EIE ++ +E + + +

Sbjct: 541 KMNQQE-ETDKTPATVKSDSKRSYEEKEWKKERQRLRRIEIEITTVQTIEENISRND 599

15 Query: 604 LMLET---NDIGKLSDLQKELESIQEEQLLMEEWENLNMRLD 643
L+ + D K+ + + E + +E L+ EWE L+ D

Sbjct: 600 LLCDEPVYQDHEKVQAIHADNEKLNQELESLLSEWEELSTEED 642

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4279> which encodes the amino acid sequence <SEQ ID 4280>. Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2042(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 473/635 (74%), Positives = 545/635 (85%), Gaps = 1/635 (0%)

Query: 9 MIILQGNKIERFSFGDVLFDNINIQVDQRDRIALVGRNGAGKSTLLKILVGEEAPTKEI 68
MIILQGNK+ERSFSGDVLFI++QVD+RDRIALVG NGAGKSTLLK+LVGEE PT GE+

35 Sbjct: 1 MIILQGNKLSRFSFGDVLFIQNISLQVDERDRIALVGPNGAGKSTLLKLLVGEEPTPTSGEV 60

Query: 69 NKKRDLISLSYLAQDSRFQSENTIFQEMLQVFDLSREVEKRLRELELQMGQVSGSDLEQLM 128
N K+DL+LSYLAQ+SRF+S+ TI++EML+VF++LR+ EKRLR++E+ M VSG L +LM

Sbjct: 61 NTKKDLTSLYLAQNSRFESDQTIYEEMLKVFALRQDEKRLQMEMDMATVSGQVLTIRM 120

40 Query: 129 KTYDILSEEFREKGGFTYESDIKAILNGFKFNSDMWEMPISLGGQNTLRALAKMLLEK 188
YD+L+E FR++GGFTYESDIKAILNGFKF+ MW+M I+ELSGGQNTLRALAKMLLEK

Sbjct: 121 TDYDLTTEHFRQCGFTYESDIKAILNGFKFDESWMQMTIAELSGGQNTLRALAKMLLEK 180

45 Query: 189 PELLVLDEPTNHLIDITIAWLENYLVNYQGALIIVSHDRYFLDKVATVTYDLTTHSLDRY 248
PELLVLDEPTNHLIDITIAWLENYL NYQGALIIVSHDRYFLDKVATVT DLT + LDY

Sbjct: 181 PELLVLDEPTNHLIDITIAWLENYLVNYQGALIIVSHDRYFLDKVATVTLDLTPNGLDY 240

Query: 249 VGNYSKFMDLKAEKIAATEKNFEKQKQKEIAKLEDFVQKNIVRASTTKRAQARRKQLEKME 308
GNYS+FM LKAEK+ EEK F+KQKQKEIAKLEDFVQ+KNIVRASTTKRAQARRKQLEK+E

50 Sbjct: 241 SGNYSRFMALKAELVAEEKQFDKQKQKEIAKLEDFVQKNIVRASTTKRAQARRKQLEKIE 300

Query: 309 RLDKPNVEQKSANMTFHAGKVSNNVLTLENAAGYEGVSLSEPIDLDVKKFDAIAIVGP 368
RLDKP +KSA+MTFHA K SGNVVL +E AAGY LSEPI++D+ K DAIA+VGP

55 Sbjct: 301 RLDKPTGGRKSAHMTFHAEPKSGNNVLRVEEAAIGYDQVLEPIVNDINKLDAIAVGP 360

Query: 369 NGIGKSTLIKSLVGQIPFIKGEAKLGANVETGYDQSQSNLTKTNTVLDELWDAFSTTPE 428
NGIGKSTLIK++GQ+P +KG+ K GANVETGYDQ+QS+LT +NTVL+ELW FSTTPE

Sbjct: 361 NGIGKSTLIKSIIGQLPLKQQLKYGANVETGYDQTSHTLSSNTVLEELWQDFSTTPE 420

60 Query: 429 VEIRNRLGAFLESGDDVKKSVSMLSGGERARLLAKLSMENNFLILDEPTNHLDDIDSKE 488
V+IRNRLGAFLESGDDVKKSV+MLSGGE+ARLLAKLSMENNFL+LDEPTNHLDDIDSKE

Sbjct: 421 VDIRNRLGAFLESGDDVKKSVAMLSGGEKARLLAKLSMENNFLVLDEPTNHLDDIDSKE 480

65 Query: 489 VLENALIEFDGTLFVSHDRYFINRVATKVLEISDKGSTLYLGDYDYLYTKKAELEELAR 548
VLENALI+FDGTLFVSHDRYFINR+ATKVLEI++ GSTLYLGDYDYLY KKAEELEELAR

-1534-

Sbjct: 481 VLENALIDFDGTLDFVSHDRYFINRLATKVLEITENGSTLYLGDYDYYLEKKAEEELAR 540

Query: 549 LNEEEVSASKTEIDVTSIDYETQKANQKEFRKITRRVVEIEARLEVLNENNINGLMLET 608
 L E E T DY+ OKANQKE R++TRR EIEARLE +E I M +

5 Sbjct: 541 LAAGETVEETKEASAT-DYQLQKANQKERRRLTRYEEIEARLETIEERIGAIQEDMHAS 599

Query: 609 NDIGKLSDLQKELESIQEEQLLMEEWENLNMRLD 643
 ND +L QKE + + +EQ LMEWE + +++

10 Sbjct: 600 NDTAQLIAWQKEWDQLDQEQEALMEEWETIAEQIE 634

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1395

A DNA sequence (GBSx1480) was identified in *S.agalactiae* <SEQ ID 4281> which encodes the amino acid sequence <SEQ ID 4282>. This protein is predicted to be thiophene degradation protein F (thdF). Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0876(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 A related GBS nucleic acid sequence <SEQ ID 9773> which encodes amino acid sequence <SEQ ID 9774> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4283> which encodes the amino acid sequence <SEQ ID 4284>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0795(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 384/458 (83%), Positives = 427/458 (92%)

40 Query: 12 MSITKEFDITIAAISTPLGEGAIGIVRISGTDALKIASKIYRGKDLSAIQSHTLNIGHIVD 71
 MSITKEFDTI AISTPLGEGAIGIVR+SGTDAL IA +++GK+L + SHT+NYGHI++
 Sbjct: 1 MSITKEFDTITAISTPLGEGAIGIVRLSGTDALAIQSVFKGNLEQVASHNTINYGHIIN 60

45 Query: 72 PDKNEILDEVMLGVMLAPKTFTRDVIEINTHGGIAVTNEILQLILRHGARMAPGEFTK 131
 P I+DEVML+ VMLAPKTFTR+V+EINTHGGIAVTNEILQL++R GARMAPGEFTK
 Sbjct: 61 PKTGTIIDVMVSVMLAPKTFTRDVIEINTHGGIAVTNEILQLILRHGARMAPGEFTK 120

50 Query: 132 RAFLNGRVDLTQAEAVMDLIRAKTDKAMDIAVKQLDGSLSKTLINNTROEILNTLAQVEVN 191
 RAFLNGRVDLTQAEAVMD+IRAKTDKAM IAVKQLDGS L LN+TROEILNTLAQVEVN
 Sbjct: 121 RAFLNGRVDLTQAEAVMDIIRAKTDKAMTIAVKQLDGSLSQLINDTROEILNTLAQVEVN 180

55 Query: 192 IDYPEYDDVEEMTTTLMREKTQEFQALMENLLRTARRGKILREGLSTAIIGRPNVGKSSS 251
 IDYPEYDDVEEMTT L+REKTQEFQ+L+E+LLRTA+RGKILREGLSTAIIGRPNVGKSSS
 Sbjct: 181 IDYPEYDDVEEMTTALLREKTQEFQSLLESLLRTAKRGKILREGLSTAIIGRPNVGKSSS 240

Query: 252 LNNLLREEKAIVTDIEGTTTRDVIEEYVNIKGVPKLVDTAGIRDTDDIVEKIGVERSKKA 311
 LNNLLRE+KAIVTDI GTTRDVIEEYVNIKGVPKLVDTAGIR+TDD+VE+IGVERSKKA
 Sbjct: 241 LNNLLREDKAIVTDIAGTTTRDVIEEYVNIKGVPKLVDTAGIRETDDLVEQIGVERSKKA 300

-1535-

Query: 312 LEEADLVLLVLSSEELTLQDRSLLELSKESNRIVLLNKTDLPQKIEVNELPKNVIPISV 371
 L+EADLVLLVLSSEELTLQDR+LL LS++SNRI+LLNKTDLP QKIE+ +LP + IPISV
 Sbjet: 301 LQEADLVLLVLSSEELTLQDRALLNLSQDSNRILLNKTDLEQKIEQLPDDYIPISV 360

Query: 372 LENENIDKIEERINDIFFDNAGMVEHDTYLSNARHISLIEKAVDSLKAVNEGLELGMPIV 431
 L N+NI+ IE+RIN +FFDNAG+VE DATYLSNARHISLIEKAV SL+AVN+GL LGMPV
 Sbjet: 361 LTQNINILIEDRINQLFFDNAGLVEQDATYLSNARHISLIEKAVQSLEAVNDGLALGMPIV 420

Query: 432 DLLQVDMTRTWELGEITGDAAPDELITQLFSQFCLGK 469
 DLLQVD+TRTWELGEITGDAAPDELITQLFSQFCLGK
 Sbjet: 421 DLLQVDLRTTWELGEITGDAAPDELITQLFSQFCLGK 458

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1396

A DNA sequence (GBSx1481) was identified in *S.agalactiae* <SEQ ID 4285> which encodes the amino acid sequence <SEQ ID 4286>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.18 Transmembrane 280 - 296 (276 - 299)
 INTEGRAL Likelihood = -4.83 Transmembrane 249 - 265 (243 - 266)

----- Final Results -----
 bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD40365 GB:AF036485 hypothetical protein [Plasmid pNZ4000]
 Identities = 88/306 (28%), Positives = 149/306 (47%), Gaps = 17/306 (5%)

Query: 1 MIVEQKFGNGFTWIN---IEAEQLRTETSEIQAKY-LDSEIITYALDDYERAFMECSHIK 56
 MI +K NG WI I AE+ T ++ +Y +D +II Y D+ E I
 Sbjet: 1 MIKPEKTINGTKWIETIQINAEERAT---LEDQYGIDEDIEYVTDNDDESTNYVYD-IN 55

Query: 57 GKEVLTIIFNTIDLKQESYYETVPMFCLSHDRLLITVTRSRNSYMBELLQKYLDNRNPDV 116
 + L I L+ Y T P L L T +S + L LD NP+V
 Sbjet: 56 EDDQLFIFLAPYALDKDALRYITQPFGLMLHKGVLFTFNQSGIPEVNTALYSALD-NPEV 114

Query: 117 -SPKKFLFAALTILTKQYFNVVSKIDREKIDILNRQLREQTNRLLAMSDLETGVSVYLLT 175
 S F+ L + + + I +++++ L++ L +T N L+++S L+ +L +
 Sbjet: 115 KSVDAFILETLFTVVVSFIPISRAITKKRNYLDKMLNRKTKNSDLVSLSYLQQTTLTFLSS 174

Query: 176 AANQNALVLEQLDVHPSQRFNSEVRKEQLS---DALIEAHQLVSMTQLNSQVLSQLSSTF 232
 A N L +LD P F +++++ D IE Q+ M ++ +QV+ ++ T
 Sbjet: 175 AVQTN---LSELDRLPKTHFGVGADQDKIDLFEEDVQIEGEQVQRMFEIETQVVDRIHTL 231

Query: 233 NNVLNNNLNENLTGLNIIISINLAIIAAITGFFGMNIPPLTESRSSWLIVIATSVLLWVI 292
 N++ NNNLN+ + L I S+ +A+ I+GF+GMN+ LPL + +W++ + SV+L V
 Sbjet: 232 NSLANNLNLDTMKFLTIWSLIMAVPTTIISGFYGMNVKLPLAGMQYAWMLTLGISVVLIVA 291

Query: 293 IAQILK 298
 + +LK
 Sbjet: 292 MLIMLK 297

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1536-

Example 1397

A DNA sequence (GBSx1482) was identified in *S.agalactiae* <SEQ ID 4287> which encodes the amino acid sequence <SEQ ID 4288>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 38
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.1437(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1398

A DNA sequence (GBSx1483) was identified in *S.agalactiae* <SEQ ID 4289> which encodes the amino acid sequence <SEQ ID 4290>. This protein is predicted to be exonuclease RexA. Analysis of this protein sequence reveals the following:

```

20  Possible site: 52
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.3165(Affirmative) < succ>
25  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9771> which encodes amino acid sequence <SEQ ID 9772> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC12966 GB:U76424 exonuclease RexA [Lactococcus lactis]
  Identities = 522/1211 (43%), Positives = 747/1211 (61%), Gaps = 73/1211 (6%)

Query: 28  KRTPEQIEAIYTFGNVLSASAGSGKTFVMVERILDKLLRGVPIDSLFISTFTVKAAGE 87
35  K TPEQ EAI++ G N+LVSASAGSGKTFVM +RI++K+ +G+ ID LFISTFT KAA E
Sbjct: 5  KLTPEQNEAIHSSGKNILVSASAGSGKTFVMAQRIVEKVKQGIEIDRLFISTFTKKAASE 64

Query: 88  LKERLEKKINESLKSASDDLLKQFLTQQLVGIQTADIGTMDAFTQKIVNQYGYTLGISPI 147
40  L+ RLE+ + ++ + + D+ LT L + ADIGTMD+FTQK+ + I P
Sbjct: 65  LRMRLERDLKKARQESSDDEEZHRLTLALQNLNADIGTMSFTQKLTKANFNVRNIDPN 124

Query: 148 FRILQDKNEQDVIIKNEVYADLFSDYMTGKNAAS-----FIKLVKNFSGNRKDSKAFREMV 202
45  FRIL D+ E D+I+ EV+ L Y++ + + F KL+KNFS +R + F+++V
Sbjct: 125 FRILADQTESDLIRQEVFEQVLVESYLSADESLNISKDKFEKLIKNSKDR-NILGFQKV 183

Query: 203 YKVYAFSQSTDNPKRWMQTVFLKGAQTYTDFEAIQDQEVSSLLNVMQT--TANQLRDLTD 260
50  Y +Y F+ +T+NP W++ FLKG +TY +++ D +NV + T +L +
Sbjct: 184 YTIYRFASATENPISWLENQFLKGFETY---KSLTDLSEDFTVNVKENLLTFFELLEAIS 240

Query: 261 QEDYKQLTAKGVPTANYKKHLKIIENL-VHNSQDFNLLYGKKGLTNLARDITNVIPSGND 319
55  ++D+ TA L I ++ V S+D L KK + +D+
Sbjct: 241 KKDFVTCTAL-----FLSIDTDIRVGSSKDEALSALKKDFSAQKQDL----- 282

Query: 320 VTVAGVKYPIFKQLHNRIIVGLKHLEVIFKYQGSESLFLELLQSFVLDLDFSEQYLQEKIQEN 379

```

-1537-

V P +L + +KH ++I KYQ ++ + LQ F++DF + YL+ K EN
 Sbjct: 283 --VGSKSKP--GELRKFDKIKHGOIEKYQNQAFETASDLQKFIDFYKTYLERKKNNEN 338
 Query: 380 AFEFSDIAHFAIQILEENHDIRQLYQDKYHEVMVDEYQDNNHTQERMLELLSNGHNRFMV 439
 AFE+SDIAHFAI+ILEEN DIR+ ++ Y E+M+DEYQD +HTQERMLELLSNGHN FMV
 Sbjct: 339 AFEYSDIAHFAIEILEENPDIRENLRHYDEIMIDEYQDTSHTQERMLELLSNGHNLFMV 398
 Query: 440 GDIKQSIYFRQADPQIFNDKYKAYQDNPSQGLIILKENFRSQSEVLDSTNSVFTHLMD 499
 GDIKQSIY FR ADP +F +KYK+Y + +LI LKENFRS+ EVL+ TN +F HLMD
 Sbjct: 399 GDIKQSIYGFRLADPGLFLEKYKSYDQAEENPNQLIRLKENFRSRGEVLNFTNDIFKHLMD 458
 Query: 500 EEVGDIYDESHQLKAGS----PRQGERHPNNKTQVLLLDTEDDIDDSOQYDISPAE 555
 E++G++ Y + L G+ P + E+ + + +T E++I+DS+ + IS E
 Sbjct: 459 EKLGEPTYGKEEALVQGNISDYPVEAEKDFYPELLLYKENTSEEEIEDSEVK---ISDGE 515
 Query: 556 AKLVAKEIIRLHKEENVFPQDITLLVSSRTRNDGILQTFDRYGIPLVTDGGEQNYLKSVE 615
 K A+EI +L E V +DI +LV S++ N+ I Y IP+V D G ++LKS+E
 Sbjct: 516 IKGAAQEIKKL-IEYGVPEKDIAILVRSKSNNNKIEDILLSYDIPVVLDEGRVDFLKSME 574
 Query: 616 VMVMLDTRLRSIDNPLNDYALVALLRSPMFGFNEDDLTRIAIQDVK-MAFYHKVKLSYHKE 674
 V++MLD LR+IDNPL D +LVA+LRSP+FGFNED+LTRI++Q + + F+ K+ LS KE
 Sbjct: 575 VLIMLDVLRADNPLYDLSLVAMLRSPFGFNEDELTRISVQGSRLDRFWDKILLSLKE 634
 Query: 675 GHSDLITPELSSKIDHFMKTFQTRWDFAKWHSLYDLIWKIYNDRFYDYVGALPKAEQR 734
 G + +LI L K+ F + F WR ++ L+WKIY + +Y+DYVGAL E R
 Sbjct: 635 GKNPELINLSLEQKLKAFNQKFTWRKLVNKIPIHRLWKIYTETYYFDYVGALKNGEMR 694
 Query: 735 QANLYALALRANQFEKTFGKLSRFIRMIDKVLNENDLADVEVALPQNAVNLMTIHKSK 794
 QANL AL++RA +E +G+KGL +F+R+I+K +E NDIA V + LPONAV +MT HKSK
 Sbjct: 695 QANLQALSVRAESYESSGYKGLFKFVRLINKFMEQNNDLASVNIKLQNAVVRVMTFHKSK 754
 Query: 795 GLEFKYVFILNIDKKFSMDVITSPLILSRNQGIGIKYVADMRHELEE-EILPAVKVSMET 853
 GLEF YVF++N+ +F+ D+ +ILSR G+G+KY+AD++ E + P V MET
 Sbjct: 755 GLEFDYVFLMNLQSRFNDRDLKEDVILSREHGLGMKYIADLKAEPDVTITDFPYALVKMET 814
 Query: 854 LPYQLNKRELRLATLSEQMRLLYVAMTRAEEKLYLVGKASQT---KWADHYDLVS-ENNH 909
 PY +NK + A LSE+MR+LYVA TRA+KKLYLVGK T + YD + E
 Sbjct: 815 FPYMVNKDLKQRAALSEEMRVLYVATRAKKLYLVGKIKDTDKKAGLELYDAATLEGKI 874
 Query: 910 LPLASRETFVTFQDWLLAVHETYKKQELFYDINFVSLEELTDHHIGMVNPSLPFNPDK- 968
 L R + FQ W+LA+ K L +N + +EL + + PD K
 Sbjct: 875 LSDKFRNSSRGFQHWILALQNA TK---LPMKLNVTYKDELETEKLEFTS-----QPDFKK 926
 Query: 969 -VENRQSEDIVRAIS--VLESVEQINQTY--KAAIELPTVTRPSQVKK-IYEPILDIEGV 1022
 VE + D + + S + E+ + +N Y +AA EL +++TPSQVKK YE L + V
 Sbjct: 927 LVESEKFDNIMSFSDIEKAQKIMNYQYPHQAATELSSIQTSPSQVKKRSYEKQLQVGEV 986
 Query: 1023 D-VMETITKTSVDFKLPDFSTSKKQDPAALGSVHELMQRIEMSSHVKMEDIQKALTEVN 1081
 V E + ++DF DF KK A +GSA H MQ + S + Q L E+
 Sbjct: 987 QPVSEFVRVKNLDFS--DFG-PKKITAEMGSATHSMQYADF-SQADLFSFQATLDEMG 1042
 Query: 1082 AETSVKAAIQIEKINYFFQETSLGKYIQEEVEHLHREAPFAMLKEDPESGKFFVVRGIID 1141
 + +K I I KI F +T G+++ E V+ +EAPF+ML+ D + E+++VRGI D
 Sbjct: 1043 FDEKIKNQIDITKILTLE-DTEFGQLSENVDKTVKEAPFSMLRTDEFAKEQYIVRGICD 1101
 Query: 1142 GYLLLENRIILFDYKTDKVFNP---LELKERYQGOMALYAEALKKSYBIEKIDKYLILLG 1198
 G++ L ++IILFDYKTD+F N E+KERY+ QM LY+EAL+K+Y + +IDKYLILLG
 Sbjct: 1102 GFVKLADKIILFDYKTDRTNVSAISEIKERYKQMNLYSEALQKAYHVNQIDKYLILLG 1161
 Query: 1199 G-KQLEVVKMD 1208
 G +++ V K+D
 Sbjct: 1162 GPRKVFVEKID 1172

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4291> which encodes the amino acid
 sequence <SEQ ID 4292>. Analysis of this protein sequence reveals the following:

Possible site: 61

-1538-

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAC12966 GB:U76424 exonuclease RexA [Lactococcus lactis]
 Identities = 478/1206 (39%), Positives = 700/1206 (57%), Gaps = 65/1206 (5%)

 Query: 40 KRTAQQIEAIYTSQONILVSASAGSGKTFVMVERILDKILRGVSIIDRLFISTFTVKAATE 99
 K T +Q EAI++SG+NILVSASAGSGKTFVM +RI++K+ +G+ IDRLFISTFT KAA+E
 Sbjct: 5 KLTPRQNEAIHSSGKNILVSASAGSGKTFVMAQRIVEKVKQIEIDRLFISTFTKKAASE 64

15 Query: 100 LRERIEKNLYSQAQTTDFQMKVYLTEQLQSLCQADIGTMDAFAQKVVSRYGYSIGISSQ 159
 LR R+E L +++D + LT LQ+L ADIGTMD+F QK+ + I
 Sbjct: 65 LRMRLELDLKKARQESSDDEEAHRLTLALQNLNADIGTMDSFTQKLTKANFNVRNVIDPN 124

20 Query: 160 FRIMQDKAEQDVLKQEVFSKLFNEFMNQKEA-----PVFRALVKNFSGNCKDTSAFRELV 214
 FRI+ D+ E D+++QEVF +L +++ E+ F L+KNFS + ++ F+++V
 Sbjct: 125 FRILADQTESDLIRQEVFEQLVESYLSADESLNISKDKFEKLIKNFPSKD-RNILGFQKV 183

25 Query: 215 YTCYSFSQSSTENPKIWLQENFLSAAKYQRLDIPDHDIELLLAMQDTANQLRDVTDME 274
 YT Y F+ +TENP WL+ FL +TY+ L D+ + D + + T +L + +
 Sbjct: 184 YTIYRFASATENPISWLENQFLKGFETYKSLTDLSE-DFTVNVKENLLTFFELLEAISKK 242

30 Query: 275 DYGQLTKAG-SRSAKYTKHLTIIEKLSDWVRDFKCLYGKAGLDRLIRDVIGLIPSGNDVT 333
 D+ T S + E LS +DF D+
 Sbjct: 243 DFVTC TALFLSIDTDIRVGSSKDEALSALKKDFSA-----QKQDLV 283

35 Query: 334 VSKVKYPVFKTLHQKQLKQFRHLETILMYQKDCFSLEQLQDFVLAFASEAYLAVKIQESAF 393
 SK K + K+K H + I YQ F + LQ F++ F + YL K E+AF
 Sbjct: 284 GSKSKPGELRKPFVDKIK---HGOLIEKYQNAFELASDLQKFIIDFYKTYLERKKENAF 340

40 Query: 394 EFS DIAHFAIKILEENTDIRQSYQQHYHEVMVDEYQDNHMQERLLTLLSNGHNRFMVGD 453
 E+S DIAHFAI+ILEEN DIR++ ++HY E+M+DEYQD +H QER+L LLSNGHN FMVGD
 Sbjct: 341 EYS DIAHFAIEILEENPDIRENLRREHYDEIMIDEYQDTSHTQERMLELLSNGHNLFMVGD 400

45 Query: 454 IKQSIYFRQADPQIFNQKFRDYQKKEQKGVILLKENFRSQSEVLNVSNVAVFSLHMD 513
 IKQSIY FR ADP +F +K++ Y + ++I LKENFRS+ EVLN +N +F HLMDE
 Sbjct: 401 IKQSIYGFRLADPGLFLEKYKSYDQAEENPNQLIRLKENFRSRGEVLNFTNDIFKHLMD 460

50 Query: 514 VGDVLYDEQHOLIAG--SHAQTVPYLDRRAQLLLYNSDKDDGNAPSDEGISFSEVTIVA 571
 +G++ Y ++ L+ G S D +LLLY + + IS E+ A
 Sbjct: 461 LGEMTYGKEALVQGNISDYPAEAKDFYPELLLYKENTSEEEIEDSEVKISDGEIKGAA 520

55 Query: 572 KEI IKIHNDKGVFPEDITLLVSSRTRNDIISHTFNQYGIPIATDGGQONYLKSVEVMV 631
 +EI KL + GV +DI +LV S++ N+ I Y IP+ D G+ ++LKS+EV++ML
 Sbjct: 521 QEIKKL-IEYGVPEKDIAILVRSKSNNNKIEDILLSYDIPVVLDEGRVDFLKSMEVLIML 579

60 Query: 632 DTLRTINNPRNDYALVALLRSPMFAFDEDDLARIALQKDNELDKDCLYDKIQRAVIGRA 691
 D LR I+NP D +LVA+LRSP+F F+ED+L RI++Q +L +DKI ++ G
 Sbjct: 580 DVLRAIDNPLYDLSLVAMLRSPLFQFNEDELTRISVQSGSRDLR---FWDKILLSLKKEGK 636

65 Query: 692 HP ELIHDTLLGKLVFLKTLKSWRRYAKLGSLYDLIWKIFNDRFYDFVASQAKAEQAQA 751
 +PELI+ +L KL F + WR+ ++ L+WKI+ + +YFD+V + E QA
 Sbjct: 637 NPELINLSLEQKLKAFNQKFTWRKLVNKKIPIHRLWLKIYTYTYFDYVGALKNGEMRQA 696

 Query: 752 NLYALALRANQFEKSGYKGLYRFIKMIDKVLLETQNDLADVEVATPKQAVNLTTHKSKGL 811
 NL AL++RA +E SGYKGL++F+++I+K +E NDLA V + P+ AV +MT HSKSKGL
 Sbjct: 697 NLQALSVRAESYESSGYKGLFKFVRLINKFMEQNNDLASVNIKLQNAVVRMTFHKSKGL 756

 Query: 812 QFPYVFILNCDKRFSMTDIHKSFILNRQHGIGIKYLADIKGLLGE-TTLNSVKVSMETLP 870
 +F YVF++N RF+ D+ + IL+R+HG+G+KY+AD+K T V MET P
 Sbjct: 757 EFDYVFLMNLQSRFNDRDLKEDVILSREHGLGMKYIADLKAEPDVTITDFPYALVKMETFP 816

-1539-

Query: 871 YQLNKQELRLATLSEEMRLLYVAMTRAEEKVYFIGK--ASKSKSQEITDPKKL-GKLLP 926
Y +NK + A LSEEMR+LYVA TRA+KK+Y +GK K E+ D L GK+L

Sbjct: 817 YMVNKDLKQRAALSEEMRVLYVAFTRAKKLYLVGKIKDTPDKAGLELYDAATLEGGKLS 876

5 Query: 927 LALREQLLTFQDWLLAIADIFSTEDLYFDVRFIEDSDLTQESVGRQLQTP---QLLNPDLL 983
R FQ W+LA+ + L + +L E + P +L+ +

Sbjct: 877 DKFRNSSRGFQHWILALQ---NATKLPMLNVYTKDELETEKLEFTSQPDFKCLVEESEK 933

10 Query: 984 KDNRSQSETIARALDMLEAVSQNLNANY--BAAIHLPTVRTPSQL-KATYEPLLEPIGVDDI 1040
DN S + ++ EA +N Y +AA L +++TPSQ+ K +YE L+ V +

Sbjct: 934 FDNIMSFSD---EIKEAQKIMNYQYPHQATELSSIQTPSQVKKRSYEKQLQVGEVQPV 989

15 Query: 1041 EKSSRSLSDFTLPHFSKKAKVEASHIGSALHQLMQVLPLSKP--INQQTLLDALRGIDSN 1098
+ R + + F K K+ A+ +GSA H MQ S+ + Q LD + G D

Sbjct: 990 SEFVR-VKNLDFSDFGPK-KITAEEMGSATHSFMQYADFSQADLFSFQATLDEM-GFD-- 1044

20 Query: 1099 EEVKTALDLKKIESFFCDTSLGQFFQTYQKHLRYEAPFAILKLDPISQEEYVLRGIIDAY 1158
E++K +D+ KI + F DT GQF +EAPF++L+ D ++E+Y++RGI D +

Sbjct: 1045 EKIKNQIDITKILTLF-DTEFGQLSENVDKTVKEAPFSMLRTDEFAKEQYIVRGICDGF 1103

25 Query: 1159 FLFDDHIVLVVDYKTDKYPQ---IELKKRYQQLELYAEALTQTYKLPVTKRYLVLMGGG 1215
D I+L DYKTD++ E+K+RY+ Q+ LY+EAL + Y + +YL+L+GG

Sbjct: 1104 VKLADKIILFDYKTDRTINVSASEIKERYKQMNLYSEALQKAYHVNQIDKYLILLGGP 1163

Query: 1216 KPEIVE 1221
+ VE

Sbjct: 1164 RKVFVE 1169

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 728/1211 (60%), Positives = 916/1211 (75%), Gaps = 5/1211 (0%)

Query: 1 MMTFKPFLNPEDIAVIQTEEKNSDKKQKRTPEQIEAIYTFGNNVLVSASAGSGKTFVMVE 60
+++F PFL+PE I +Q E+ D+ QKRT +QIEAIYT G N+LVSASAGSGKTFVMVE

35 Sbjct: 13 VISFAPFLSPEAIKHLQENRRCRDQSQRKTAQQIEAIYTSQNILVSASAGSGKTFVMVE 72

Query: 61 RILDKLLRGVPIDSLFISTFTVKAAGELKERLEKKINESLSAESDDLKQFLTQQLVGIQ 120
RILDK+LRGV ID LFISTFTVKA EL+ER+E K+ + +K +LT+QL +

Sbjct: 73 RILDKILRGVSIDRLFISTFTVKAATELRERIEKNLYSQIAQTTFQMKVYLTEQLQSLC 132

40 Query: 121 TADIGTMDAFTQKIVNQYGYTLGISPIFRILQDKNEQDVIKNEVYADLFSDYMTGKNAAS 180
ADIGTMDAF QK+V++YGY++GIS FRI+QDK EQDV+K EV++ LF+++M K A

Sbjct: 133 QADIGTMDAFAQKVVSRYGYSIGISSQFRIMQDKAEQDVLKQEVFSKLFNEFMNQKEAPV 192

45 Query: 181 FIKLVKNFSGNRKDSKAFREVMYKVYAFSQSTDNPKRWMQTVFLKGAQTYTDFEAIQDE 240
F LVKNFSGN KD+ AFRE+VY Y+FSQST+NPK W+Q FL A+TY E IPD +

Sbjct: 193 FRALVKNFSGNCKDTSARELVYTCYSFSQSTENPKIWLQENFLSAAKTYQRLIEDIPDHD 252

50 Query: 241 VSSLLNMVQTTANQLRDLTDQEDYKQLTAKGVPTANYKKHLKIIENLVHWSQDFNLLYGK 300
+ LL MQ TANQLRD+TD EDY QLT G +A Y KHL IIE L W +DF LYGK

Sbjct: 253 IELLLAMQDTANQLRDVTDMDYQQLTKAGSRSKAYTKHLTIIEKLSDWVRDFKCLYGK 312

55 Query: 301 KGLTNLARDITNVIPSGNDVTAVGVKYPFKQLHNRIVGLKHLEVIFKYQGSESLFLELL 360
GL L RD+T +IPSGNDVTV+ VKYP+FK LH ++ +HLE I YQ + LLE L

Sbjct: 313 AGLDRILRDVTGLIPSGNDVTVSVKYPVFKTLHQKLQFRHLETILMYQKDCFSLEQL 372

60 Query: 361 QSFVLDPSEQYLQEKIQENAFEFSDIAHFAIQILEENHDIRQLYQDKYHEVMVDEYQDNN 420
Q FVL PSE YL KIQE+AFEFSDIAHFAI+ILEEN DIRQ YQ YHEVMVDEYQDNN

Sbjct: 373 QDFVLAFSEAYLAVKIQESAFEFSDIAHFAIKILEENTDIRQSYQQHYHEVMVDEYQDNN 432

65 Query: 421 HTQERMLELLSNGHNRFMVGDIKQSIYFRQADPQIFNDKYKAYQDNPSQCKLIILKENF 480
H QER+L LLSNGHNRFMVGDIKQSIYFRQADPQIFN K++ YQ P QGK+I+LKENF

Sbjct: 433 HMQERILLTLLSNGHNRFMVGDIKQSIYFRQADPQIFNQKFRDYQKKPEQGGKVIILKENF 492

Query: 481 RSQSEVLDSSTNSVFTHLMDVEVGDILYDESHQLKAGSPRQQRHPNPKTQVLLLDTEDD 540
RSQSEVL+ +N+VF+HLMDE VGD+LYDE HQL AGS Q + + + Q+LL ++D+DD

Sbjct: 493 RSQSEVLNVSNVAFSHLMDESVDVLYDEQHQLIAGSHAQTVPYLDRRAQLLLYNSDKDD 552

-1540-

Query: 541 IDSDSQYDISPAEAKLVAKIIRLHKEENVFQDITLLVSSRTRNDGILQTFDRYGIP 600
 ++ S IS +E +VAKEII+LH ++ VPF+DITLLVSSRTRND I TF++YGIP
 Sbjct: 553 -GNAPSDSEGISFSEVTIVAKEIILHNDKGVFPEDITLLVSSRTRNDIISHTFNQYGIP 611

5 Query: 601 LVTGGEQNYLKSVEVMVMDLTLRSIDNPLNDYALVALLRSPMFGFNEDDLTRIAIQD-- 658
 + TDGG+QNYLKSVEVMVMDLTLR+I+NP NDYALVALLRSPMF F+EDDL RIA+Q
 Sbjct: 612 IATDGGQNYLKSVEVMVMDLTLRTINNPRNDYALVALLRSPMFAFDEDDLARIALQKDN 671

10 Query: 659 --VKMAFYHKVLSYHKEGHSDLTPELSSKIDHFMKTFQTWRDFAKWHSLYDLIWKIY 716
 K Y K++ + G H +LI L K++ F+KT ++WR +AK SLYDLIWKI+
 Sbjct: 672 ELDKDCLYDKIQRAVIGRGAHPELIHDTLLGKLNVLKTLKSWRRYAKLGSYDLIWKIF 731

15 Query: 717 NDRFYDYVGALEPKAEQOQANLYALALRANQFEKTFKGLSRFIRMIDKVLNENNDLADV 776
 NDRFY+D+V + KAEQ QANLYALALRANQFEK+G+KGL RFI+MIDKVLN +NDLADV
 Sbjct: 732 NDRFYDFVASQAKAEQAQANLYALALRANQFEKSGYKGLYRFIKMIDKVLNQNLDADV 791

20 Query: 777 EVALPQNAVNLMTIHKSGLEFKYVFILNIDKKFSMDITSPLILSRNQGIGIKYVADMR 836
 EVA P+ AVNLMTIHKSGGL+F YVFILN DK+FSM DI IL+R GIGIKY+AD++
 Sbjct: 792 EVATPKQAVNLMTIHKSGGLQFPYVFILNCDKRFSDTIHKSFILNRQHGIGIKYLADIK 851

25 Query: 837 HELEEEILPAVKVSMETLPYQLNKRRLATLSEQMRLLYVAMTRAEEKLYLVGKASQTK 896
 L E L +VKVSMETLPYQLNK+ELRLATLSE+MRLLYVAMTRAEEK+Y +GKAS++K
 Sbjct: 852 GLLGETTLNSVKVSMETLPYQLNKQELRLATLSEEMRLLYVAMTRAEEKVYFIGKASKSK 911

30 Query: 897 WADHYDLVSENNHPLASRETFVTFQDWLLAVHETKYKQELFYDINFVSLLEELTDHHIGM 956
 + D LPLA RE +TFQDWLLA+ + + ++L++D+ F+ +LT +G
 Sbjct: 912 SQEITDPKKGKLLPLALREQLLTFQDWLLAIADIFSTEDLYFDVRFIEDSDLTQESVGR 971

35 Query: 957 VNPSLPFNPDKVENRQSEDIVRAISVLESVEQINQTYKAAIELPTVRTPSQVKKIYEPI 1016
 + NPD+ +NRQSE I RA+ +LE+V Q+N Y+AAI LPTVRTPSQ+K YEP+
 Sbjct: 972 LQTPQLNPDLDKDNROSETIARALDMLEAVSQLNANYEAAIHLPTVRTPSQLKATYEP 1031

40 Query: 1017 LDIEGVDMETITKTSVDFKLPDFSTSKKQDPAALGSVAVHELMQRIEMSSHVKMEDIQKA 1076
 L+ GVD++E +++ DF LP FS K + + +GSA+H+LMQ + +S + + + A
 Sbjct: 1032 LEPIGVDIIEKSSRSLSDFTLPHFSKKAKVEASHIGSALHQLMQVLPLSKPINQOTLLDA 1091

45 Query: 1077 LTEVNAETSVKAAIQIEKINYFFQETSLGKYIQEEVEHLHREAPFAMLKEDPESGEKFFV 1136
 L +++ VK A+ ++KI FF +TSLG++ Q +HL+REAPFA+LK DP S E++V+
 Sbjct: 1092 LRIGDSNEEVKTALDLKKIESFFCDTSLGQFFQTYQKHLREAPFAILKLDPISQEEYVL 1151

50 Query: 1137 RGIIDGYLLENRIILFDYKTDKFNPLELKERYQQOMALYAEALKKSYBIEKIDKYLIL 1196
 RGIID Y L ++ I+L DYKTDK+ P+ELK+RYQ Q+ LYAEAL ++Y++ +YL+L
 Sbjct: 1152 RGIIDAYFLFDDHIVLDYKTDKYKQPIELKKRYQQLELYAEALTQTYKLPVTKRYLV 1211

55 Query: 1197 LGGKQLEVVKM 1207
 +GG + E+V++
 Sbjct: 1212 MGGGKPEIVEV 1222

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1399

A DNA sequence (GBSx1484) was identified in *S.agalactiae* <SEQ ID 4293> which encodes the amino acid sequence <SEQ ID 4294>. This protein is predicted to be exonuclease RexB. Analysis of this protein sequence reveals the following:

55 Possible site: 23
 >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0660(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1541-

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC12965 GB:U76424 exonuclease RxB [Lactococcus lactis]

Identities = 363/1093 (33%), Positives = 604/1093 (55%), Gaps = 67/1093 (6%)

5 Query: 1 MKLLYTDINHDMTEILVNQAAHAAEAGWRIFVIAPNSLSFEKERAVLENLPQ---EASFA 57
M++LYT+I D+TE L+ A E +++YI P+S+SFEKE+ +LE L + A F
Sbjct: 1 MEILYTEITQDLTEGLLEIALBEELEKNRKVYIVPSSMSFEKEKEILERLAKGSDTAVFD 60

10 Query: 58 ITITRFAQLARYFTLNQP-NQKESLNDIGLAMIFYRALASFEDGQLKVFGRKQDASFIS 116
+ +TRF QL YF + K L +GL+M+F R L SF+ ++ ++ L+ A F+
Sbjct: 61 LLVTRFKQLPYFYDKREKATMKTGLTGVLSMLFRRVLRFSFKKDEIPLYFSIQDSAGFLE 120

15 Query: 117 QLVDLYKELQTANLSILELKYLHSPEKFEDLLAIFLVVSDLLREGEYDNQSKIAFFTEQV 176
L+ L EL TANLS+ L ++ + +LA F + EY N S+ FT ++
Sbjct: 121 MLIQLRABELLTANLSVENLPDNPKNQELKKILAKFEAELSV----EYANYSEFGDFINRL 176

20 Query: 177 RSGQLDVLKNTILVDGFRSAAEEELIKSLSSRCQEIIGAYASQKAYKANFTNGNI 236
G+ D LK+ +I+DG+TRFSABEE I+S+ + ++G Y+ + + A + I
Sbjct: 177 VDGEFDQQLKDVITIIIDGYTRFSABEEELFIESIQEKVARFVVGTYSDENSLTAG--SETI 234

25 Query: 237 YSAGVDFLRYLATTFQTKPEFILSKWESKSGFEMISK-----NIEGKHDFTNSSHILDDT 291
Y + T F+ K L K S + E+ SK +++ + T+ L
Sbjct: 235 YVGTSQMI----TRFRNKFVVELRKIASSAVNEVYSKLTRILDLSRFVITDEKIELKAE 290

30 Query: 292 AKDCITIWECINQKDEVEHVARAIRQKLYQGYRYKDILVLLGDVDVSYKLQLSKIFEQYDI 351
+ IWE NQK E+E VA+ IRQK+ QG +KD VL+GD +Y++ L ++F+ Y+I
Sbjct: 291 DEKYFRWEAENQKVEIERVAKEIRQKIIQGAFFKDFTVLVGDPAAYETLKEVFDLYEI 350

35 Query: 352 PYYFGKAETMAAHLVHFMDSLSRIKRYRFRADVLNLFKTGIYGEISQDD--LDYFEAY 409
P+++ + E+M+ HPLV F +SL IK+ +R +DV+NL K+ +Y + + D+ +DYFE Y
Sbjct: 351 PFFYAQBESMSQHPLVIFVESLFAIKKNYRTDDVVNLKSKVYTDANLDEEVIDYFEY 410

40 Query: 410 ISYADIKGPKKFFTDVVGAKKFDLGRNLNIRQSL---TPLESFV-KTKKQDGIKTLNQ 465
+ I G KKF +F+ ++ + +N +R+ LL +PL+ F+ +K+ G K ++
Sbjct: 411 VQKYKISGRKKFTEEFIE-SFESQIELVNEMREKLLGSESPLQVFLGNNRKKTKKKWVSD 469

45 Query: 466 FMFFLTQVGLSDNLSRLVQMS-ENEQE---KHQEVWKTFTDILEQFQTFGQEKLNLE 521
L + N++ +NE + KH++VW+ L +F +F EKL E
Sbjct: 470 LQGLENGNVMTNMNAYFSAELQNEHQMAKHEQVWQMLISTLNEFLAVFSDEKLSVE 529

50 Query: 522 FLSLLNSGMMQAEYRMVPATVDVVTVKSYDLVEPHSNQFVYALGMTQSHFPKIAQNKSLI 581
FL +L +G+ A+YR +PA VDVV VK Y+LVEP +N+++YA+G+++Q++FP+I +N +L+
Sbjct: 530 FLDILLAGLKNAYRQIPANVDVNVKDYELVEPKTNKYIYAIGLSQTNFPRIKKNSTLL 589

55 Query: 582 SDIERQLINDANDTDGHFDIMTQENLKNHFAALSFLNAKQELVLTIPQLNESEDQMS 641
SD ER IN D + + + N +KN F LSL N+AK+ LVL++PQ++ + + S
Sbjct: 590 SDEERLEINQTTDENQFIEQLNVANYQKNQFTVLSLINSAKESLVLSMPQIMANEQGEFS 649

60 Query: 642 P-YLVELRDIGVPFNHKGK-RQLKEEADNIGNYKALLSRVVDLYRSAIDKEMTKEE-QTF 698
P + + L+D K + +L E ++IGN +++++ + + R ++ E T E+ + F
Sbjct: 650 PVFQLFLKDADEKILQKIQGVNLFESLEHIGNSRSVIAMIGQIERELVESEETSEDKRVF 709

65 Query: 699 WSVAVRYLRQLTSKGIEIPIITDSTLDTVTSSDVMTRRFPEDDPLKLSSSALTTFFYNNQ 758
WS R L + + + +DTV ++ D + + + D + S S+ FYN +
Sbjct: 710 WSSIFRILVKSNAFDQKILLDLAKDIDTVNLAPDTLEQIY--GDKIYASVSSFERFYNCE 767

70 Query: 759 YKYFLQYVLGLEEQDSIHPDMRHGTYLHRVFEILMKNQGI--ESFEKLSAINKTNQE 816
Y+YFL+ L LE ++I + + G + H VFE +MK + E+F+EKL + + ++
Sbjct: 768 YQYFLENTLSLETENIDINSKIVGNFFHEVFKEVMKETDLSAENFDEKLTLLVQLQEVDKN 827

75 Query: 817 DVFKSLYSEDAESRYSLEILEDIARATATILR----QDSQMTVESE-----EERFELM 865
+ +++DA +R++ LE+I R TAT+L+ D T+ +E E
Sbjct: 828 --YSRYFTQDATARFTWSNLEBIVRQTATVLKATVSTDELKTLTLESSFGLPKSELGNFS 885

80 Query: 866 IDNTIKINGIIDRIDRLSDGSLGVVDYKSSAQKFDIQKFYNGLSPLVTVYIDAISRDKV 925
+D+ I + G IDR+D+LS LG +DYKSSA F +Q+ Y+GLS Q +TY+D I K+
Sbjct: 886 VDD-IYLRGRIDRLDQLSTDYLGADYKSSAHSFKLQEAAYDGLSLQFMFTYLDVI---KQA 941

-1542-

Query: 926 EQKPPIFGAMYILHMQEPRQDLSKIKNLDDLVTKNHQALTYKGLFSEAEKEFLANGKYHL- 984
 I+GA+YL + +LS+I L ++ +++ Y+GL E E + G ++
 Sbjct: 942 FPNQKIWGALYLYQFKNQPINLSEINQLSEIANILKESMRVEGLVLEDAAEQI-KGIENIA 1000

Query: 985 --KDSLYSETEIAILOAHNQSLYKKASETIKSGKFLINPYTEDAKTVGD-----Q 1033
 K ++Y+E E L N+ Y+ A + +K GK INP + ++ +D
 Sbjct: 1001 LKKTNIYNEEFEQLLKLNEEHYRAAGQRLKKGKIAINPIMKRSEIDQSGNVRGCRYCP 1060

Query: 1034 FKSITGFEADRH 1046
 KSI FEA+ HM
 Sbjct: 1061 LKSICRFEANIHM 1073

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4295> which encodes the amino acid
 sequence <SEQ ID 4296>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1891(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 546/1075 (50%), Positives = 758/1075 (69%), Gaps = 11/1075 (1%)

Query: 1 MKLLYTDINHDMTIELVNQAAHAAEAGWRIFYIAPNSLSFEKERAVLENLPQEASFAITI 60
 MKL+YT++++ MTEILVN+A AA+ G+R+FYIAPNSLSFEKER VL LP+ +F+I +
 Sbjct: 1 MKLIYTEMSSYMTIELVNEARKAADQGYRVFYIAPNSLSFEKEREVLTLPERGTFSIIV 60

Query: 61 TRFAQLARYFTLNQPNQKESLNDIGLAMIFYRALASFEDGQLKVFGRKQDASFISQLVD 120
 TRF Q++RYFT+ K+ L+D LAMIFYRAL + L +GRL+ ++ FI QLV+
 Sbjct: 61 TRFVQMSRYFTVESSPSKQHLDDTTLAMIFYRALMQLKPEDLPYGRLONNVVFIEQLVE 120

Query: 121 LYKELQTANLSILELKYLSPEKFDLLAIFLVVSDLLREGEYDNQSKIAFFTEQVRSQ 180
 LYKEL+ A LS+ +L L P+K EDL+ I + ++ +Y+ S + F ++ G
 Sbjct: 121 LYKELKNAQLSVHDLTGLDHPQKQEDLIKIIELAETIMIQDYNQDSPLOSFARAIKGL 180

Query: 181 LDVDLKNLTILVDGFTFRSAEEALIKSLSSRCQEIIIGAYASQKAYKANFTNGNIYSAG 240
 L+ L T++++DGF+RFSABE+ L+ L++ CQE+IIG+Y SQKAY+ +F GNIY A
 Sbjct: 181 LNNQLSKTVVVIDGFSRFSABEEDYLLSLNNNCQEVIIIGSVVSQKAYQKSFIKGNIYEAS 240

Query: 241 VDFLRYLATTFQTKPEFILSKWESKSGFEMISKNIIEGKHDFTNSSHILDDTAKDCITWE 300
 + FL+ LA + KP F S K F +++ E HDF+ L + D ++W+
 Sbjct: 241 LHFLQDLAQYHIKPVFATSNQVFKPAFSRLTQLFEATHDFSQVDWQLQKSDLDHFLWQ 300

Query: 301 CINQKDEVEHVARAIRQKLYQGYRYKDILVLLGDVDSYKLQLSKIFEQYDIPYFGKAET 360
 C +QK+E+EHVA++IRQKLY+GYRYKDILVLLGD+D+Y+LQ+ IF++++IPY GKA
 Sbjct: 301 CHHQKEEIEHVAKSIRQKLYEGYRYKDILVLLGDMDAYQLQIGPIFDKFEIPYVLGKAEP 360

Query: 361 MAAHPLVHFMDLSLSRIKRYRFRAEDVLNLFKTIYGEISQDDLDYFEAYISYADIKGPKK 420
 MAAHPLV F++SL R +RY +R ED+LN+ K+G++G D+D FE Y +ADIKG K
 Sbjct: 361 MAAHPLVQFIESLERSQRYNWRREDILNMLKSGLFGCFDDSDIDRFEEYTQFADIKGFTK 420

Query: 421 FFTDFVV-GAKKFDLGRINTIRQSLTLPLESFVTKKQDGIKTLNQFMFFLTQVGLSDNL 479
 F F + +++ L LN +RQ ++ PL+ K++KQ G +++ + FL ++ L++N+
 Sbjct: 421 FSKPFTINSSRQYPLDFLNEMRQDIVLPLQELFKSOKQLGASLVDKLILFLKKIRLAENM 480

Query: 480 SRLVGQMSENEQEKHQEVWKTFTDILEQFQTIQFQEKLNLDLFLSLNSGMMQAEYRMVP 539
 L S+ E EK++EVWK FTDIL F IFQEK L + L+L+ +GM A+YR+VP
 Sbjct: 481 QGLA--QSQLEVEKNEEVWKRTDILTFSHHIFGQEKRLSDCLALIKTGMKSAQYRVVP 538

Query: 540 ATVDVVTVKSYDLVPHSNQFVYALGMTQSHFPKIAQNKSLISDIERQLINDANDTDGHF 599
 AT+DVVT+KSYDLV+PHS FVYA+G+TQSHFPK + L+SD ER IN+ + HF
 Sbjct: 539 ATLDVVTIKSYDLVQPHSKPFVYATGLTQSHFPKQIHSGLLSDQERARINEIRNY-RHF 597

-1543-

5 Query: 600 DIMTQENLKKNHFAALSLFNAAQELVLTIPQLLNESEDQMSPYLVELRDIGVFPNHKGR 659
DI + EN KKNH ALSLFNAA +ELVL++ ++NE+ D +SPYL EL + G+P KG+
Sbjct: 598 DIASAENSKKNHQTALSLFNAAATKELVLSVSTVINETFDLSPYLKELINFGLPLLDKGG 657

10 Query: 660 QSLKEEADNIGNYKALLSRVVDLYRSAIDKEMTKEEQTFWSVAVRYLRRLTSKGIEIPI 719
L + +IGNYKALLS+++ + R + EM+ +++ FW+V +RYLR+QL + +E+P
Sbjct: 658 NYLSYDNSDIGNYKALLSQIILAINRQDL-IEMSDQDKMFWTVVLRYLKQLRKQQLLELPT 716

15 Query: 780 RHHGTYLHRVFEILMKNQIGIESFEEKLNSAINKTNQEDVFKSLYSEDAESRYSLEILEDI 839
R HG YLHRVFE LMK+ E F+ KL AI TNQE F+ +Y ++AE+ YSL ILEDI
Sbjct: 777 RIHGQYLRVFEILMKDHTQEPFDNKLKQAIYHTNQESFFQQVYQDNAAEYSLAILEDI 836

20 Query: 840 ARATATILRQDSQMTVESEERFELMIDNTIKINGIIDRIDRLSDGSLGVVDYKSSAQKF 899
R+TA IL+ + + V +E+ F+L + N I ++GIIDRID+LSDGSLG+VDYKSSA +F
Sbjct: 837 VRSTAPILQLNQNIQVIDQEKNFQDLMGNEILVHGIIDRIDQLSDGSLGIVDYKSSANQF 896

25 Query: 900 DIQKFYNGLSPLQVITYIDAISR--DKEVEQKPPIFGAMYLMQEPQDLSKIKNLDD-LV 956
DI FYNGLSPLQ+TY+ A+ + ++ Q +FGAMYLH+Q+P+ DL K +D+ LV
Sbjct: 897 DIGTFYNGLSPLQMTYLAALKQIAPHDINQ---LFGAMYLHLQDPKLDLVTFKQIDNTLV 953

30 Query: 1017 FLINPYTEDAKTVGDQFQSITGFEADRHMARARALYKLPAGEKRGFLTLMQQE 1071
FLINPYT D KTV GDQ K+IT FEAD M +AR L LPAKEK++ FLTLM++E
Sbjct: 1014 FLINPYTSDGKTVDGDLKAITRFEADLDMGQARRLVTLPAKEKKECFLTLMRKE 1068

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1400

A DNA sequence (GBSx1485) was identified in *S.agalactiae* <SEQ ID 4297> which encodes the amino acid sequence <SEQ ID 4298>. Analysis of this protein sequence reveals the following:

40 Possible site: 31
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -7.80 Transmembrane 51 - 67 (44 - 69)

45 ----- Final Results -----
bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8799> which encodes amino acid sequence <SEQ ID 8800> was also identified. Analysis of this protein sequence reveals the following:

50 Lipop: Possible site: -1 Crend: 10
McG: Discrim Score: -20.62
GvH: Signal Score (-7.5): -6.25
Possible site: 31
>>> Seems to have no N-terminal signal sequence

55 ALOM program count: 1 value: -7.80 threshold: 0.0
INTEGRAL Likelihood = -7.80 Transmembrane 47 - 63 (40 - 65)
PERIPHERAL Likelihood = 3.34 26
modified ALOM score: 2.06

60 *** Reasoning Step: 3

-1544-

----- Final Results -----

bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC75528 GB:AE000334 orf, hypothetical protein [Escherichia coli K12]
 Identities = 138/297 (46%), Positives = 193/297 (64%), Gaps = 16/297 (5%)

10 Query: 5 MKIDDLRKSDNVEDRRSSSGGSFSSGSGLPILQLLLLRGSKTKLVVLIILLLLG--GG 62
 M+ R+SDNVEDRR+SSGG S GG G + S K L++L++L+ G G
 Sbjct: 1 MRWQGRRESNVEDRRNSSGGP-SMGGPGFRL-----PSGKGLLILVIVLVAGYGV 52

15 Query: 63 GLTSIFNDSSSPSSYSQSNVRSVDNSATREQIDFVNKVLGSTEDFWSQEFQTQGFQNYK 122
 LT + ++++S + D +A F + +L +TED W Q+F+ G Y+
 Sbjct: 53 DLTGLMTGQPVSQQQSTRSISPNEDEAAK-----FTSVILATTEDTWGQQFEKMG-KTYQ 106

20 Query: 123 EPKLVLYTNSIQTCGIGESASGPFYCSADKKIYLDISFYNELSHKYGATGDFAMAYVIA 182
 +PKLV+Y +TGCG G+S GPFYC AD +Y+D+SFY+++ K GA GDFA YVIA
 Sbjct: 107 QPKLVMYRGMTRTTCGAGQSIMGPFYCPADGTIVIDLFSFYDDMKDKLGADGDFAGQGVVIA 166

25 Query: 183 HEVGHHIQTELIGIMDKYNRMHGLTKKEANALNVRLELQADYYAGVWAHYIRGNLLEQG 242
 HEVGHH+Q LGI K +++ T+ E N L+VR+ELQAD +AGVW H ++ + +LE G
 Sbjct: 167 HEVGHHVQKLLGIEPKVRQLQONATQAEVNRLSVMELQADCFAGVWGHSMQQQGVLETG 226

Query: 243 DFEEAMNAHAHVGDDTLQKETYGKLVDPDSFTHGTAEQRQRWPNKGFQYGDQHGDTF 299
 D EEA+NAA A+GDD LQ+++ G++VPDSFTHGT++QR WF +GF GD +TF
 Sbjct: 227 DLEELNAAQAIGDRLQQSQGRVVPDSFTHGTSSQQRYSWFKRGFDSGDPAQCNTF 283

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4299> which encodes the amino acid sequence <SEQ ID 4300>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -6.42 Transmembrane 48 - 64 (41 - 67)

35

----- Final Results -----

bacterial membrane --- Certainty=0.3569(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40

The protein has homology with the following sequences in the databases:

>GP:AAC75528 GB:AE000334 orf, hypothetical protein [Escherichia coli]
 Identities = 143/301 (47%), Positives = 195/301 (64%), Gaps = 21/301 (6%)

45 Query: 1 MKTDDLRESQQVEDRRGQSSG-SFGGGGLGGGLLLQLLFSRGGWTKLVILLILLVLMG-- 57
 M+ RES VEDRR S G S GG G +L +GG L++L+++LV G
 Sbjct: 1 MRWQGRRESNVEDRRNSSGGPSMGGPGF-----RLPSGKGG----LILLIVLVVAGYY 50

50 Query: 58 GGGLSGVLGGKPSSTNNAYQSSQVTRTNGDKASQEQVSFVSKVFASTEDYWTKTFREKG 117
 G L+G++ G+P S QS++ N D+A++ F S + A+TED W + F + G
 Sbjct: 51 GVDLTGLMTGQPVSQQ-----QSTRSISPNEDEAAK-----FTSVILATTEDTWGQQFEKMG 102

55 Query: 118 LTYHKPTLVLYTGATQTACGRGQASSGPFYCPGDQKVYLDISFYNELSTKYGAKGDFAMA 177
 TY +P LV+Y G T+T CG GQ+ GPFYCP D VY+D+SFY+++ K GA GDFA
 Sbjct: 103 KTYQQPKLVMYRGMTRTTCGAGQSIMGPFYCPADGTIVIDLFSFYDDMKDKLGADGDFAGQ 162

60 Query: 178 YVIAHEVGHHIQNELGIMDNYASARQKSKAKANQLNVKLELQADYYAGAWANYVQQGQL 237
 YVIAHEVGHH+Q LGI +Q ++A+ N+L+V++ELQAD +AG W + +Q QG+
 Sbjct: 163 YVIAHEVGHHVQKLLGIEPKVRQLQONATQAEVNRLSVMELQADCFAGVWGHSMQQQGV 222

Query: 238 LEKGDIEEAMAAAHAVGDDTLQEEYGRTPVDPDSFTHGTSKQRQRWFDGRGYQYGDFFHGDTF 298
 LE GD+EEA+ AA A+GDD LQ+++ GR VPDSFTHGT+QR WF RG+ GD +TF
 Sbjct: 223 LETGDLEELNAAQAIGDRLQQSQGRVVPDSFTHGTSSQQRYSWFKRGFDSGDPAQCNTF 283

-1545-

An alignment of the GAS and GBS proteins is shown below.

Identities = 191/303 (63%), Positives = 241/303 (79%), Gaps = 5/303 (1%)

```

5  Query: 5  MKIDDLRKSDNVEDRRSSSGSFSSGG-SGLPILQLLLLLRGSWKTKLVVLIILLGSGG 63
    MK DDLR+S VEDRR S GSF GG G +LQLL RG WKTKLV+L++LL++GGGG
    Sbjct: 1  MKTDDLRESQQVEDRRSQSSGSGGGLGGGLLQLLFSRGGWTKLVILLLLVMGGG 60

10  Query: 64  LTSIFN---DSSSPSSYQSONVSRSDNSATREQIDFVNKVLGSTEDFWSQEFQTQGGFN 120
    L+ +      S++ ++YQS V+R+ + A++EQ+ FV+KV STED+W++ F+ +G
    Sbjct: 61  LSGVLGGKPSSTNNAYQSSQVTRTNGDKASQEQVSFVSKVFASTEDYWTKTFRKGL-T 119

15  Query: 121  YKEPKLVLYTNSIQTCGIGESASGPFYCSADKKIYLDISFYNELSHKYGATGDFAMAYV 180
    Y +P LVLYT + QT CG G+++SGPFYC D+K+YLDISFYNELS KYGA GDFAMAYV
    Sbjct: 120 YHKPTLVLYTGATQTACGRGQASSGPFYCPGDQKVYLDISFYNELSTKYGAKGDFAMAYV 179

20  Query: 181  IAHEVGHHIQTELGIMDKYNRMHGLTKKEANALNVRLQLADYYAGVWAHYIRGKNLLE 240
    IAHEVGHHIQ ELGIMD Y   R G +K +AN LNV+LELQADYYAG WA+Y++G+ LLE
    Sbjct: 180 IAHEVGHHIQNELGIMDNYASARQGKSKAKANQLNVKLELQADYYAGAWANYVQGQGLLE 239

25  Query: 241  QGDFEEMANAAHVGDDTLQKETYGKLVDPDSFTHGTAEQRQRWFNKGFGQYGDIGHGDTFS 300
    +GD EEAM AAHAVGDDTLQ+ETYG+ VPDSFTHGT++QRQRWF++G+QYGD +HGDTFS
    Sbjct: 240 KGDIEEAMAAAHVGDDTLQEEETYGRTPDSFTHGTSKQRQRWFDRCGYQYGD FEHGDTFS 299

    Query: 301  VEH 303
    + +
    Sbjct: 300  IPY 302
  
```

SEQ ID 8800 (GBS404) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 3; MW 62kDa).

GBS404-GST was purified as shown in Figure 218, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1401

A DNA sequence (GBSx1486) was identified in *S.galactiae* <SEQ ID 4301> which encodes the amino acid sequence <SEQ ID 4302>. This protein is predicted to be phenylalanyl-tRNA synthetase beta chain (pheT). Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have no N-terminal signal sequence
  
```

```

40  ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2617(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  
```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB14823 GB:Z99118 phenylalanyl-tRNA synthetase (beta subunit)
[Bacillus subtilis]
Identities = 376/805 (46%), Positives = 523/805 (64%), Gaps = 6/805 (0%)
  
```

```

50  Query: 1  MLVSYKWLKELVDVD-VTTAELAEKMSTTGIEVEGVETPAEGLSKLVVGHIVSCEDVPDT 59
    M VSYKWL++ VD+ + A LAEK++ GIEVEG+E EG+ +V+GH++ E P+
    Sbjct: 1  MFVSYKWLEDYVDLKGMDPAVLAEKITRAGIEVEGIEYKGEIGKGVVIGHVLERQHPNA 60

55  Query: 60  H-LHLCQVDTGDDDELQVVCAGPNVKTGINVIVAPGARIADNYKIKKGKIRGMESLGM 118
    L+ C VD G + Q++CGAPNV G V VA GA + N+KIKK K+RG ES GMI
    Sbjct: 61  DKLNKCLVDIGAEAPVQIICGAPNVDKGQKQVAVATVGAVLPGNFKIKKAKLRGBESNGMI 120
  
```

-1546-

Query: 119 CSLQELGLSESIIPKEFSDGIQILPEGAIPGDSIFS YLDDDEIIELSTTPNRADALSMR 178
 CSLQELG+ ++ KE+++GI + P A G + L LDD I+EL +TPNRADA++M
 Sbjct: 121 CSLQELGIESKLVAKEYARGIFVFPNDAETGSDALAALQLDDAILELGLTPNRADAMNML 180

5 Query: 179 GVAHEVAAIYGKKVHFEEKNLIEEAERAADKISVVIESDKVLS-YSARIVKNVTVPSPQ 237
 GVA+EVA AI +V + + +E+A+D ISV IE + Y+A+I+KNVT+APSP
 Sbjct: 181 GVAYEVAAILDTEVKLPQTDYPAASEQASDYISVKIEDQEANPLYTAKIKNVTIAPSFL 240

10 Query: 238 WLQNKLMNAGIRPINNVVDVTNYVLLTYGQPMHAFDFDKFDGTTIVARNAENGEKLITLD 297
 W+Q KLMNAGIRP NNVD+TN+VLL YGQP+HAFD+D+F +V R A E ++TLD
 Sbjct: 241 WMQTKLMNAGIRPHNNVDITNFVLLYEGQPLHAFD YDRFGSKEVVVRKAAENEMIVTLD 300

15 Query: 298 GEERDLIADDLVIAVNDQPVALAGVMGGQSTEIGSSSKTVVLEAAVFNGETSIRKTSGRIN 357
 +ER L AD LVI + A+AGVMGG +E+ +KT++LEAA FNG +RK S L
 Sbjct: 301 DQERKLSADHLVITNGTKAQAVAGVMGGAESEVQEDTKITLLEAAVFNQKVRKASKDLG 360

20 Query: 358 LRSESSSRFEKGINYDVTSEAMDFAAAMLQELAGGQVLSGGVTEGVLPTPEVEVSTTLGY 417
 LRSESS RFEKGI+ V A + AA ++ AGG+VL+G V E L E + +
 Sbjct: 361 LRSESSVRFEKGIDPARVRLAAERAAQLIHLYAGGEVLACTVEEDHLTTEANNIHVSADK 420

25 Query: 418 VNTRLGTELTYTDIEEVFEKLGFASGSEVKFTVLVPRRRWDIAIQADLVEEIRIYGYE 477
 V++ LG ++ ++ ++++LGF + ++ V VP RR DI I+ DL+EE AR+YGY+
 Sbjct: 421 VSSVLGLTISKEELISYKRLGFTVGEADDLLVVTVPSRRGDITTEEDLIEEARLYGYD 480

30 Query: 478 KLPTTLPEAGATAGELTSMQRLRRRVRTVAEGAGLSEIITYALTTPKEAVQFSTQATNIT 537
 +P+TLPE T G LT Q RR+VR EGAGLS+ ITY+LT +KA F+ + + T
 Sbjct: 481 NIPSTLPETAGTTGGLTPYQAKRRKVRREFEGAGLSQAITYSLTNEKKATAFAIEKSLNT 540

35 Query: 538 ELMWPMTVDRSALRQNVVSGMLDTIAYNVARKNSNLAVYEIGKVFEQTGNPKEDLPTEVE 597
 L PM+ +RS LR ++V +LD+++YN+AR+ ++A+YE+G VF ++ P E E
 Sbjct: 541 VLALPMSEERSILRHSVLPNLLDSVSYNLARQTDSEVALYEVGSVF--LTKEEDTKPVETE 598

40 Query: 598 TTFALTGLVEEKDFQTKSKPVDFFYAKGIVEALFIXLK-LDVTFVAQKGLASMHGRTA 656
 A+TGL ++ +Q + KPVDFF KGIVE L KL LD Q +HPGRTA
 Sbjct: 599 RVAGAVTGLWRKQLWQGEKKPVDFFVVKGIVEGLLDKLNVLDSIEFVQSERKQLHPGRTA 658

45 Query: 657 TILLDGKEIGFVGQVHPQIAKQYDIPETYVABINLSTIESQMNOALIFEDITKYPVSRD 716
 ILL+G IGF+GQVHP K+ DI ETYV E++L + + L++ I KYPVSRD
 Sbjct: 659 NILNGSLIGFIGQVHPSLEKELDIKETYVFELDLHALLAETAPLVYTAIPKYPVSTRD 718

50 Query: 717 IALLAESVSHHDIVSAIETSGVKRLTAIKLFDVYAGNNIAEGYKSMAYSLTFQNPNDNL 776
 IAL+ ++V+ + S I+ +G K L + +FDVY G ++ EG KS+A+SL + NP L
 Sbjct: 719 IALVTDKTVTSGQLESVIKEAGGKLLKEVTVFVYEGEHMEEGKKSVAFSLQYVNPQTL 778

55 Query: 777 TDEEVAKYMEKITKSLVEKVNARIR 801
 T+EEV K K+ K+L + A +R
 Sbjct: 779 TEEVTKAHSKVLKALEDTYQAVLR 803

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4303> which encodes the amino acid
 50 sequence <SEQ ID 4304>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1283(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

60 Identities = 595/801 (74%), Positives = 687/801 (85%)
 Query: 1 MLVSYKWLKELVDVDTTAEALAEKMSTTGIEVEGVETPAEGLSKLVVGHIVSCEDVPDTH 60
 MLVSYKWLKELVD+DVT A LAEKMSTTGIEVEG+E PA+GLSKLVVGH++SCEDVP+TH
 Sbjct: 6 MLVSYKWLKELVDIDVTPAALAEKMSTTGIEVEGIEVPADGLSKLVVGHVLSCEDVPETH 65

-1547-

Query: 61 LHLCCQVDTGDDDELRCVVCAPNVKTGINVIVAVPGARIADNYKIKKGKIRGMESLGMICS 120
 LHLCCQVDTGD+ RQ+VCGAPNVK GI VIVAVPGARIADNYKIKKGKIRGMESLGMICS
 Sbjct: 66 LHLCCQVDTGDETPRQIVCGAPNVKAGIKVIVAVPGARIADNYKIKKGKIRGMESLGMICS 125

5 Query: 121 LQELGLSESIIPKEFSDBGIQLPEGAIPGDSIFS YLDLDDDEIIELSITPNRADALSMRGV 180
 LQELGLS+SIIPKEFSDBGIQLPE A+PGD+IF YLDLDD IIELSITPNRADALSMRGV
 Sbjct: 126 LQELGLSDSIIPKEFSDBGIQLPEEAVPGDAIFKYLDLDDHIIELSITPNRADALSMRGV 185

10 Query: 181 AHEVAAIYKKVHFEEKNLIEEAADKISVVIESDKVLSYSARIVKNVTVAPSPQWLQ 240
 AHEVAAIYKG V F +KNL E + ++ I V I SD VL+Y++R+V+NV V PSPQWLQ
 Sbjct: 186 AHEVAAIYKSVSFPQKNLQESDKATSEAIEVAIASDNVLTYSRVVENVKVPSPQWLQ 245

15 Query: 241 NKLMMAGIRPINNVVDVTNYVLLTYGQPMHAFDFDKFDGTTIVARNAENGEKLITLDGEE 300
 N LMNAGIRPINNVVDVTNYVLL +GQPMHAFD+DKF+ IVAR A GE L+TLDGE+
 Sbjct: 246 NLLMAGIRPINNVVDVTNYVLLYFGQPMHAFDYDKFEDHKIVARAARQGESLVTLDGEEK 305

20 Query: 301 RDLIADDLVIAVNDQPVLAGVMGGQSTEIGSSSKTVVLEAAVFNSTIRKTSGRNLNRS 360
 RDL +DLVI V D+PVALAGVMGGQ+TEI ++S+TVVLEAAVF+G SIRKTSGRNLNRS
 Sbjct: 306 RDLITTEDLVITVADKPVALAGVMGGQATEIDANSQTVVLEAAVFDGKSIRKTSGRNLNRS 365

25 Query: 361 ESSSRFEKGINYDVTSEAMDFAAAMLQELAGGQVLSGQVTEGVLPTEPVEVSTITLGYVNT 420
 ESSSRFEKG+NY TV EA+DFAAAMLQELA GQVLSG V G LPTEPVEVST+L YVN
 Sbjct: 366 ESSSRFEKGVNYATVLEALDFAAAMLQELAEQGQVLSGHVQAGQLPTEPVEVSTSLDYVNV 425

30 Query: 421 RLGTELTYTDIEEVFEKLGFAISGSEVKFTVLVPRRRWDIAIQADLVEEIIARIYGYEKL 480
 RLGTELT+ DI+ +F++LGF ++G E FTV VPRRRWD++I ADLVEEIIARIYGY+KLP
 Sbjct: 426 RLGTELTFADIQRIQIFDQLGFGLTGDETSFTVAVPRRRWDVSIPADLVEEIIARIYGYDKLP 485

35 Query: 481 TTLPEAGATAGELTSMQRLRRRVRTVAEGAGLSEIITYALTTPPEKAVQFSTQATNITELM 540
 TTLPEAG TA ELT Q LRR+VR +AEG GL+EII+YALTTPPEKAV+F+ +++TELM
 Sbjct: 486 TTLPEAGTAABELTPTQALRRKVRGLAEGLGLTEIISYALTTPPEKAVEFAVAPSHLTEL 545

40 Query: 541 WPMTVDRSALRQNVVSGMLDTIAYNVARKNSNLAVVEIGKVFEQTNPKEDLPTEVETFT 600
 WPM+V+RSALRQNV+VSGMLDT+AYNVARK SNLA+YEIGK+FEQ NPKEDLP EV F
 Sbjct: 546 WPMVSVERSALRQNMVSGMLDTVAYNVARKQSNLALYEIGKIFEQANPKEDLPNEVNHFA 605

45 Query: 601 FALTGLVEEKDFQTKSKPVDFFYAKGIVBALFIKLDVTFVAQKGLASMHPGRTATILL 660
 FA+ GLV +KDFQT+++ VDF++AKG ++ LF L L V +V K LA+MHPGRTA ILL
 Sbjct: 606 FAICGLVAQKDFQTAQAVDFYHAKGNLDTLFLANLNLKVQVVPKDLANMHPGRTALILL 665

50 Query: 661 DGKEIGFVGQVHPQTAKQYDIPETVVAEINLSTIESQMNOALIFEDITKYPVSVDIAL 720
 D + IGFGVQVHP TAK Y IPETVVAE++++ +E+ + F +ITK+P+++RD+ALL
 Sbjct: 666 DEQVIGFVGQVHPGTAKAYSIPETVVAELDMAALEALPSDQTFAEITKFPAMTRDVALL 725

55 Query: 721 LAESVSHHDIVSAIETSGVKRLTAIKLFDVYAGNNIAEGYKSMAYSLTFQNPNDNLTD 780
 L VSH IV+AIE++GVKRLT+IKLFDVY G I G KSMAYSLTFQNPNDNLTD
 Sbjct: 726 LDREVSHQAIVTAIESAGVKRLTSIKLFDVYEGATTQAGKSMAYSLTFQNPNDNLTD 785

Query: 781 VAKYMEKITKSLVEKVNAEIR 801
 VAKYMEKITK+L E+V AE+R
 Sbjct: 786 VAKYMEKITKALTEQVGAEV 806

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 1402

A DNA sequence (GBSx1487) was identified in *S. agalactiae* <SEQ ID 4305> which encodes the amino acid sequence <SEQ ID 4306>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

60

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0653(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1548-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9769> which encodes amino acid sequence <SEQ ID 9770> was also identified.

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15205 GB:Z99120 transcriptional regulator [Bacillus subtilis]
Identities = 60/169 (35%), Positives = 100/169 (58%)

Query: 17 ITFKKVGLDNVNILQNIATETFRQTFSDNSEEQLQAFFNESYTLFVLKSEITHAESDTY 76
+ KK +++ LQ ++IETF TF NS E ++A+ ++ L+ E+++ S +
10 Sbjet: 3 VKMKKCSREDLQTLQQLSIETFNDFKEQNSPENMKAYLESANFTEQLEKELSNMSSQFF 62

Query: 77 FVYLDLTVGLVYLVNWSQQTEKDLDFEIQRIYLLDAYQGGIGKATFEFALDLAYS 136
F+Y D ++ GY+KVN Q+E+ ++ EI+RIY+ +++Q G+GK A+++A +
15 Sbjet: 63 FIYFDHBIAGYVKVNIDDAQSEEMGAESLEIERIYIKNSFQKHGLGKHLNKAIEIALER 122

Query: 137 GLDWAUWLVGVEFNHKAQAFYAKYGFKEFSEHQFSVGDKVDTDWLLRKSL 185
WLGWVE N A AFY K GF + H F +GD+ TD ++ K+L
20 Sbjet: 123 NKKNIWLVGVEKNENAIAFYKKMGFVQTGAHSFYMGEDEQTDLIMAKTL 171

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1403

25 A DNA sequence (GBSx1488) was identified in *S.agalactiae* <SEQ ID 4307> which encodes the amino acid sequence <SEQ ID 4308>. This protein is predicted to be phenylalanyl-tRNA synthetase (alpha subunit) (pheS). Analysis of this protein sequence reveals the following:

Possible site: 45
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3937 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9339> which encodes amino acid sequence <SEQ ID 9340> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14824 GB:Z99118 phenylalanyl-tRNA synthetase (alpha subunit)
[Bacillus subtilis]
Identities = 209/338 (61%), Positives = 270/338 (79%), Gaps = 2/338 (0%)

40 Query: 1 MKISTQEKLEKEM-TGNHTKELQDLRVQVLGKKGSLTEILLKGLKDLSDNLRPVVGKQVNEV 59
+K QE L+++ + K + D+RVQ LGKKG +TE+L+G+ LS + RP +G NEV
45 Sbjet: 5 LKQLEQEALEQVEAASSLKVVNDIRVQYLGKKGPITEVLRGMGKLSAEERPKMGALANEV 64

Query: 60 RDILTKAFFEEQAKVVEAAKIQAELESESVDTLPGRQMTLGHHRVLTQTSEEIEDIFLGM 119
R+ + A ++ + +E +++ +L +++DVTLP + +G RH LT EEIED+F+GM
50 Sbjet: 65 RERIANAIADKNEKLEEEEMKQKLAGQTIDVTLPGNPVAVGGRHPLTVVIEEIEDLFIGM 124

Query: 120 GFQVVDGFEVEKDYYNFERMNLKPHARDMQDTFYITEEILLRTHSTSPVQARTMDQHD 179
G+ V +G EVE DYYNFE +NLPK+HPARDMQD+FYITEE L+RT TSPVQ RTN++H+
Sbjet: 125 GYTVEEGFEVETDYYNFESLNLKPHARDMQDSFYITEETLMRTQTSPVQTRTMKHE- 183

55 Query: 180 SKGPLKMISPGRVFRRTDDATHSHQFHQIEGLVVGGENISMGDLKGTLLQLISQKMFGAER 239

-1549-

KGP+K+I PG+V+RRD DDATHSHQF QIEGLVV +NISM DLKGT+L+++KMFG +R
 Sbjct: 184 GKGPVKIICPGKVYRRDNDATHSHQFMQIEGLVVDKNISMSDLKGTLELVAKMFMGQDR 243
 Query: 240 KIRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCKQTGWIEILGAGMVHPSVLEMSGIDSE 299
 +IRLRPS+FPFTEPSVEVDV+CFKCGG GC+VCK TGWIEILGAGMVHP+VL+M+G D +
 Sbjct: 244 EIRLRPSFFPFTEPSVEVDVTCFKCGGNGCSVCKGTGWIEILGAGMVHPNVLKMGFDPK 303
 Query: 300 KYSGFAFGLGQERIAMLRYGINDIRGFYQGDVRFTDQF 337
 +Y GFAFG+G ERIAML+YGI+DIR FY DVRF QF
 Sbjct: 304 EYQGFAGMGVERIAMLYKYGIDDIRHFYTNDRVFISQF 341

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4309> which encodes the amino acid sequence <SEQ ID 4310>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2806(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 305/337 (90%), Positives = 327/337 (96%)

Query: 1 MKISTQEKLEKEMTGNHTKELQDLRVQVLGKKGSLTELLKGLKDLNDLRFPVVGKQVNEVR 60
 +K T E L+ +TGNHTKELQDLRV VLGKKGSLTELLKGLKDLNDLRFPVVGKQVNEVR
 Sbjct: 36 LKTKTLETLQSLTGNHTKELQDLRVAVLGKKGSLTELLKGLKDLNDLRFPVVGKQVNEVR 95
 Query: 61 DILTKAFEEQAKVVEAAKIQALQLESESVDTLPGRQMTLGHHRVLTQTSEEIEDIFLGMG 120
 D+LTKAFEEQAK+VEAAKIQAL++ES+DVTLPGRQMTLGHHRVLTQTSEEIEDIFLGMG
 Sbjct: 96 DLLTKAFEEQAKIVEAAKIQALDABSIDVTLPGRQMTLGHHRVLTQTSEEIEDIFLGMG 155
 Query: 121 FQVVDGFEVEKDYYNFERMNLPKDHPARDMQDTFYITEEILLRTHSTSPVQARTMDQHDFFS 180
 FQ+VDGFEVEKDYYNFERMNLPKDHPARDMQDTFYITEEILLRTHSTSPVQART+DQHDFFS
 Sbjct: 156 FQIVDGFVEVEKDYYNFERMNLPKDHPARDMQDTFYITEEILLRTHSTSPVQARTLDQHDFFS 215
 Query: 181 KGPLKMISPGRVFRRDTHDATHSHQFHQIEGLVVGGENISMGDLKGTLLQLISQKMFGAERK 240
 KGPLKM+SPGRVFRRDTHDATHSHQFHQIEGLVVG+NISMGDLKGT+L++I +KMFG ER
 Sbjct: 216 KGPLKMVSPGRVFRRDTHDATHSHQFHQIEGLVVGKNISMGDLKGTLEMIKKMFGDERS 275
 Query: 241 IRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCKQTGWIEILGAGMVHPSVLEMSGIDSEK 300
 IRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCK+TGWIEILGAGMVHPSVLEMSG+D+++
 Sbjct: 276 IRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCKTGWIEILGAGMVHPSVLEMSGVDAKE 335
 Query: 301 YSGFAFGLGQERIAMLRYGINDIRGFYQGDVRFTDQF 337
 YSGFAFGLGQERIAMLRYGINDIRGFYQGD RF++QF
 Sbjct: 336 YSGFAFGLGQERIAMLRYGINDIRGFYQGDQRFSEQF 372

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1404

A DNA sequence (GBSx1489) was identified in *S.galactiae* <SEQ ID 4311> which encodes the amino acid sequence <SEQ ID 4312>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2834(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1550-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1405

A DNA sequence (GBSx1490) was identified in *S.agalactiae* <SEQ ID 4313> which encodes the amino acid sequence <SEQ ID 4314>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2762(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1406

A DNA sequence (GBSx1491) was identified in *S.agalactiae* <SEQ ID 4315> which encodes the amino acid sequence <SEQ ID 4316>. This protein is predicted to be DNA-entry nuclease. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8801> which encodes amino acid sequence <SEQ ID 8802> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1   Crend: 5
McG: Discrim Score:      10.13
GvH: Signal Score (-7.5): -5.07
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
ALOM program  count: 1 value: -6.79 threshold: 0.0
  INTEGRAL    Likelihood = -6.79   Transmembrane    8 - 24 ( 6 - 27)
  PERIPHERAL  Likelihood = 6.26    258
modified ALOM score: 1.86

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```


-1551-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
Identities = 154/232 (66%), Positives = 180/232 (77%), Gaps = 1/232 (0%)

5   Query: 41  KNVSGTPSRELSSESVLTSNVKKQLGTNIAWNQSGAFIINQNKTDLNAKVSSAPYAINI 100
      K S PS+ L+ESVLT VK Q+ ++ WN SGAFI+N NKT+L+AKVSS PYA N+ K
      Sbjct: 43  KQASEAPSQALAESVLTDAVKSQIKGSLEWNGSGAFIVNGNKTNLDAKVSSKPYADNNTK 102

10  Query: 101 KVVNQIVPTKANALLTKATRQYRNREETGNGRTYWKPGWHQINGLKGSYNHAVDRGHLI 160
      V + VPT ANALL+KATRQY+NR+ETGNG T W P GWHQ+ LKGSY HAVDRGHL+
      Sbjct: 103 TVGKETVPTVANALLSKATRQYKRNKRTGNGSTSWIPPGWHQVKNLKGSYTHAVDRGHLL 162

      Query: 161 GYALVGSLSRGFDASTSNPKNIATQAAWANOANSNOSTGQNYETLVRKALDRHKT VRYRV 220
      GYAL+G L GFDASTSNPKNIA Q AWAQA + STGQNYE+ VRKALD++K VRYRV
15  Sbjct: 163 GYALIGGLDGFDASTSNPKNIAVQTAWANOQAQAEYSTGQNYYESKVRKALDQNKRVRYRV 222

      Query: 221 TLIY-DRDNLSSGSHIEAKSSDGSLEFNVFIPNVQSGLLFDYATGKVKQTK 271
      TL Y ++L+ S S IEAKSSDG LEFNV +PNVQ GL DY TG+V T+
20  Sbjct: 223 TLYYASNEDLVPSASQIEAKSSDGELEFNVLPNVQKGLQLDYRTGEVTVTQ 274
```

There is also homology to SEQ IDs 368 and 1302.

SEQ ID 8802 (GBS285) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 6; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 7; MW 57.5kDa).

25 GBS285-GST was purified as shown in Figure 208 (lane 7) and Figure 225 (lane 8).

GBS658 was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 134 (lane 8 & 9; MW 27kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 1407

A DNA sequence (GBSx1492) was identified in *S.agalactiae* <SEQ ID 4317> which encodes the amino acid sequence <SEQ ID 4318>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.

35  ----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

45 Example 1408

A DNA sequence (GBSx1493) was identified in *S.agalactiae* <SEQ ID 4319> which encodes the amino acid sequence <SEQ ID 4320>. This protein is predicted to be UDP-N-acetylglucosamine (murA). Analysis of this protein sequence reveals the following:

-1552-

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1814(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9767> which encodes amino acid sequence <SEQ ID 9768>
 10 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15693 GB:Z99122 UDP-N-acetylglucosamine
 1-carboxyvinyltransferase [Bacillus subtilis]
 Identities = 248/423 (58%), Positives = 323/423 (75%), Gaps = 5/423 (1%)

15 Query: 5 MDKIIVEGGQTQLQGQVVEGAKNAVLPLLAATILPSQGKTLTNNVPILSDVFTMNNVVR 64
 M+KIIV GGQ +L G V +EGAKNAVL++AA++L S+ K+++ +VP LSDV+T+N V+R
 Sbjct: 1 MEKIIVRGGQ-KLNGTVKVEGAKNAVLPIAASLLASEEKSVICDVPTLSDVYITINEVLR 59

20 Query: 65 GLDIQVDFNCDKKEILVDASGDILDVAPYEFVSQMRASIVVLGPILARNGHAKVSMPPGC 124
 L V F + E+ V+AS + AP+E+V +MRAS++V+GP+LAR GHA+V++PGGC
 Sbjct: 60 HLGADVHF--ENNEVTVNASYALQTEAPFEYVRKMRASVLVMGPLLARTGHARVALPGGC 117

25 Query: 125 TIGSRPIDLHLKGLEAMGATITQNGGDITAE-CLKGANIYMDFFPSVGATQNLMAATL 183
 IGSRPID HLKG EAMGA I G I A+ + +L+GA IY+DFPSVGAT+NL+MAA L
 Sbjct: 118 AIGSRPIDQHLKGFEAMGAIEKVGNGFIEAEVKGRLOGAKIYLDFFPSVGATENLIMAAAL 177

30 Query: 184 ASGTTTIENAAREPEIVDLAQLINKMGAKVKAGTETLTITIGVDALHGTEDVDVQDRIEA 243
 A GTTT+EN A+EPEIVDLA +N MG K++GAGT T+ I GV+ LHG +H ++ DRIEA
 Sbjct: 178 AEGTTTLENVAKEPEIVDLANYINGMGGKIRGAGTGTIKIEGVEKLHGKHHIIPDRIEA 237

35 Query: 244 GTFMVAAAMTSGNVLVKDAIWEHNRPLISKLMEMGVEVSEEDGIRVKADTKKLKPVTVK 303
 GTFMVAAA+T GNVLVK A+ EH LI+K+ EMGV + +E +G+RV K+LKP+ +K
 Sbjct: 238 GTFMVAAAITEGNVLVKGAVEHLTSLIAKMEEMGVTIKDEGEGLRV-IGPKELKPIDIK 296

40 Query: 364 GGRALQGAFFVMSTDLRASAALILAGMVAQGGQTVVGQLTHLDRGYQFHEKLAALGANIKR 423
 G LQGA V +TDLRA AALILAG+VA+G T V +L HLDRGY FH+KLAALGA+I+R
 Sbjct: 357 GPVQLQGAEEVAATDLRAGAALILAGLVAEGHTRVTELKHLDRGYVDFHQKLAALGADIER 416

45 Query: 424 VSE 426
 V++
 Sbjct: 417 VND 419

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4321> which encodes the amino acid
 sequence <SEQ ID 4322>. Analysis of this protein sequence reveals the following:

50 Possible site: 39
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.03 Transmembrane 377 - 393 (376 - 394)

----- Final Results -----

55 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

60 >GP:CAB15693 GB:Z99122 UDP-N-acetylglucosamine
 1-carboxyvinyltransferase [Bacillus subtilis]
 Identities = 248/423 (58%), Positives = 318/423 (74%), Gaps = 5/423 (1%)

-1553-

5 Query: 1 VDKIIIEGGQTRLEGEVVEIEGAKNAVLPLLAASILPSKGKTLIRNVPILSDVFTMNNVVR 60
 ++KII+ GGQ +L G V +EGAKNAVL++AAS+L S+ K+++ +VP LSDV+T+N V+R
 Sbjct: 1 MEKIIIVRGGQ-KLNGTVKVEGAKNAVLPIAASLLASEEKSVICDVPILSDVYTINEVLR 59

10 Query: 61 GLDIRVDFNEAANEITVDASGHILDEAPYEYVSQMRASIVVLGPILARNGHAKVSMPPGC 120
 L V F NE+TV+AS + EAP+EYV +MRAS++V+GP+LAR GHA+V+++PGGC
 Sbjct: 60 HLGADVHFEN--NEVTVNASYALQTEAPFEYVRKMRASVLVMGPILLARTGHARVALPGGC 117

15 Query: 121 TIGSRPINLHLKGLEAMGATITQKGGDITAQAD-RLQGAMIYMDFPSVGATQNLMMMAATL 179
 IGSRP I+ HLKG EAMGA I G I A+ RLQGA IY+DFPSVGAT+NL+MAA L
 Sbjct: 118 AIGSRPIDQHLKGFEAMGAETKVGNGFIEAEVKGRLQCAKIYLDFPSVGATENLIMAAAL 177

20 Query: 180 ADGVTTIENAAREPEIVDLAQFLNKMGARIRGAGTETLTITGVTHLRGVEHDVVQDRIEA 239
 A+G TT+EN A+EPEIVDLA ++N MG +IRGAGT T+ I GV L GV+H ++ DRIEA
 Sbjct: 178 AEGTTTLENVAKEPEIVDLANYINGMGGKIRGAGTGTIKIEGVEKLHGKHHIIPDRIEA 237

25 Query: 240 GTFMVAAAMTSGNVLIRDAVWEHNRPLISKLMEMGVSVTEEEYGIRVQANTPKLKPVTVK 299
 GTFMVAAA+T GNVL++ AV EH LI+K+ EMGV++ +E G+RV +LKP+ +K
 Sbjct: 238 GTFMVAAAITEGNVLKGAPEHLTSLIAKMEEMGVTIKDEGEGLRV-IGPKELKPIDIK 296

30 Query: 300 TLPHPGFPTDMAQFTALMAVNGESTMVETVFENRFQHLEEMRRMGLQSEILRETAMIH 359
 T+PHPGFPTDMQ+Q AL+ +G S + ETVFENRF H EE RRM +I + +I+
 Sbjct: 297 TMPHPGFPTDMQSQMALLLRASGTSMITETVFENRFMHAEFRMRNGDIKIEGRSVIIN 356

35 Query: 360 GGRQLQGAPVMSTDLRASAALILTGIVAQGVTVNNLVHLDRGYQFHEKLAKLGATISR 419
 G QLQGA V +TDLRA AALIL G+VA+G T V L HLDRGY FH+KLA LGA I R
 Sbjct: 357 GPVQLQGAEVAATDLRAGAALILAGLVAEGHTRVTELKHLDRGYVDFHQKLAALGADIER 416

40 Query: 420 SSE 422
 ++
 Sbjct: 417 VND 419

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 363/422 (86%), Positives = 391/422 (92%)

40 Query: 5 MDKIIIEGGQTLQGGQVVEIEGAKNAVLPLLAATILPSQGKTLTNPILSDVFTMNNVVR 64
 +DKII+EGGQ+L+G+VVEIEGAKNAVLPLLA+ILPS+GKT+L NVPILSDVFTMNNVVR
 Sbjct: 1 VDKIIIEGGQTRLEGEVVEIEGAKNAVLPLLAASILPSKGKTLIRNVPILSDVFTMNNVVR 60

45 Query: 65 GLDIQVDFNCDKKEILVDASGDILDVAPYEFVSQMRASIVVLGPILARNGHAKVSMPPGC 124
 GLDI+VDFN EI VDASG ILD APYE+VSQMRASIVVLGPILARNGHAKVSMPPGC
 Sbjct: 61 GLDIRVDFNEAANEITVDASGHILDEAPYEYVSQMRASIVVLGPILARNGHAKVSMPPGC 120

50 Query: 125 TIGSRPIDLHLKGLEAMGATITQNGGDITAQAEKLGANIYMDFPSVGATQNLMMMAATLA 184
 TIGSRPI+LHLKGLEAMGATITQ GGDITQA++L+GA IYMDFPSVGATQNLMMMAATLA
 Sbjct: 121 TIGSRPINLHLKGLEAMGATITQKGGDITAQADRLQGAMIYMDFPSVGATQNLMMMAATLA 180

55 Query: 185 SGTTTIENAAREPEIVDLAQLLNKMGAKEVKGAGTETLTIIIGVDALHGTEHDVVQDRIEAG 244
 G TTIENAAREPEIVDLAQ LNKMGAAA+GAGTETLTI GV L G EHDVVQDRIEAG
 Sbjct: 181 DGVTTIENAAREPEIVDLAQFLNKMGARIRGAGTETLTITGVTHLRGVEHDVVQDRIEAG 240

60 Query: 245 TFMVAAAMTSGNVLVKDAIWEHNRPLISKLMEMGVSEVEDGIRVKADTKKLKPVTVK 304
 TFMVAAAMTSGNVL++DA+WEHNRPLISKLMEMGV V+EEE GIRV+A+T KLKPVTVK
 Sbjct: 241 TFMVAAAMTSGNVLIRDAVWEHNRPLISKLMEMGVSVTEEEYGIRVQANTPKLKPVTVK 300

65 Query: 305 LPHPGFPTDMAQFTALMAVNGESTMIETVFENRFQHLEEMRRMGLQTEILRDTAMIH 364
 LPHPGFPTDMAQFTALMAVNGESTM+ETVFENRFQHLEEMRRMGLQ+EILR+TAMIH
 Sbjct: 301 LPHPGFPTDMAQFTALMAVNGESTMVETVFENRFQHLEEMRRMGLQSEILRETAMIH 360

Query: 365 GRALQGAPVMSTDLRASAALILAGMVAQGGTVVQGLTHLDRGYQFHEKLAALGANIKRVSE 426
 GR LQGAPVMSTDLRASAALIL G+VAQG T+V L HLDRGYQFHEKLA LGA I RSSE
 Sbjct: 361 GRQLQGAPVMSTDLRASAALILTGIVAQGVTVNNLVHLDRGYQFHEKLAKLGATISRSE 422

65 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1554-

Example 1409

A DNA sequence (GBSx1494) was identified in *S.agalactiae* <SEQ ID 4323> which encodes the amino acid sequence <SEQ ID 4324>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2096(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA23756 GB:AB009314 proton-translocating ATPase, epsilon

subunit [Streptococcus bovis]

Identities = 102/138 (73%), Positives = 121/138 (86%), Gaps = 1/138 (0%)

Query: 1 MAQLTVQVVTPDGIRYDHASLITVTRTPDGEMGILPGHINLIAPLNHVHQMKINRSHQEG- 59

M +TVQVVTPDGIRYDHHA+ I+V+TPDGEMGILP HINLIAPL VH+MKI+R+

Sbjct: 1 MTFMTVQVVTPDGIRYDHANFISVKTTPDGEMGILPEHINLIAPLTVHEMKIHRITDDPNH 60

Query: 60 VDWVAVNGGIIEVNEDQVTIVADSAERARDIDLNRAERAKERAERALEKAQTTQNIIDEMR 119

VDWVA+NGGIIE+ ++ VTIVADSAER RDID++RAERAK RAER LE+AQ+T +IDE+R

Sbjct: 61 VDWVAINGGIIEIKDNLVTIVADSAERERDIDVSRAERAKIRAERKLEQAQSTHDIDEVR 120

Query: 120 RAEVALRRRAINRISVGKK 137

RA+VALRRA+NRISVG K

Sbjct: 121 RAQVALRRALNRISVGK 138

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4325> which encodes the amino acid sequence <SEQ ID 4326>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2539(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 100/138 (72%), Positives = 119/138 (85%), Gaps = 1/138 (0%)

Query: 1 MAQLTVQVVTPDGIRYDHASLITVTRTPDGEMGILPGHINLIAPLNHVHQMKINRSHQ-EG 59

M Q+TVQVVTPDGI+YDHHA I+V TPDGEMGILP HINLIAPL VH+MKI R + E

Sbjct: 1 MTQMTVQVVTPDGIRYDHAKFISVTTTPDGEMGILPNHINLIAPLQVHEMKIRRGGEDEK 60

Query: 60 VDWVAVNGGIIEVNEDQVTIVADSAERARDIDLNRAERAKERAERALEKAQTTQNIIDEMR 119

VDWVA+NGGIIE+ ++ VTIVADSAER RDID++RAERAK RAER + +A+TT NIDE+R

Sbjct: 61 VDWIAINGGIIEIKDNLVTIVADSAERDRDIDVSRAERAKLRAEREIAQAETTHNIDEVR 120

Query: 120 RAEVALRRRAINRISVGKK 137

RA+VALRRA+NRIV K

Sbjct: 121 RAKVALRRALNRINVSCK 138

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1555-

Example 1410

A DNA sequence (GBSx1495) was identified in *S. agalactiae* <SEQ ID 4327> which encodes the amino acid sequence <SEQ ID 4328>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 60
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
10     bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein is similar to the beta subunit of the S. mutans ATPase:

      >GP:AAD13383 GB:U31170 ATPase, beta subunit [Streptococcus mutans]
      Identities = 435/466 (93%), Positives = 455/466 (97%)

15     Query: 1  MSSGKIAQVVGPPVVDVVFASGDKLPEINNALLIVYKNGDKSQKVVLEVALELGDGLVRTIA 60
      MS+GKIAQVVGPPVVDV FA+ DKLPEINNALL+VYK+GDKSQ++VLEVALELGDGLVRTIA
      Sbjct: 1  MSTGKIAQVVGPPVVDVAFATDDKLPEINNALLVYKDGDKSQRIVLEVALELGDGLVRTIA 60

20     Query: 61  MESTDGLTRGLEVLDYTGRAISVPVKGDTLGRVFNVLGDAIDLEPPFAEDAERQPIHKKAP 120
      MESTDGLTRGLEV DTGRAISVPVKG+TLGRVFNVLGD IDL++PFAEDAERQPIHKKAP
      Sbjct: 61  MESTDGLTRGLEVFDYTGRAISVPVKGDTLGRVFNVLGDTIDLDKPFADAERQPIHKKAP 120

25     Query: 121 SFDELSTSSSEILETGKIVIDLLAPYLKGGKVLFGGAGVGKTVLIQELIHNIQAQEHGGIS 180
      SFD+LSTS+EILETGKIVIDLLAPYLKGGKVLFGGAGVGKTVLIQELIHNIQAQEHGGIS
      Sbjct: 121 SFDDLSTSTEILETGKIVIDLLAPYLKGGKVLFGGAGVGKTVLIQELIHNIQAQEHGGIS 180

30     Query: 181 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE 240
      VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE
      Sbjct: 181 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE 240

35     Query: 241 QGDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYPQTLATEMGQLQERITSTKKGSVTSI 300
      QGDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYPQTLATEMGQLQERITSTKKGSVTSI
      Sbjct: 241 QGDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYPQTLATEMGQLQERITSTKKGSVTSI 300

40     Query: 301 QAIYVPADDDYTDPAAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRALTPETVGDDEH 360
      QAIYVPADDDYTDPAAPATAFAHLDSTTNLER+LTQMGIYPAVDPLASSSRAL+PEIVG EH
      Sbjct: 301 QAIYVPADDDYTDPAAPATAFAHLDSTTNLERRLTQMGIYPAVDPLASSSRALSPEIVGQEH 360

45     Query: 361 YEVATEVQRVLRQRYRELQDIIAILGMDELSDEEKTIVGRARRIQFFLSQNFNVAETFTGQ 420
      Y+VATEVQ VLRQRYRELQDIIAILGMDELSDEEKTIVGRARRIQFFLSQNFNVAE FTGQ
      Sbjct: 361 YDVATEVQHVLRQRYRELQDIIAILGMDELSDEEKTIVGRARRIQFFLSQNFNVAEQFTGQ 420

50     Query: 421 EGSYVPVEETVRGFKELDKGHDQIPEDAFRMVGGIEDVIAKAEKM 466
      EGSYVPV ETVRGFKEL+GK+D++PEDAFR VG IEDV+ KA+KM
      Sbjct: 421 EGSYVPVAETVRGFKELDKGHDQIPEDAFRSGAIEDVVEKAKKM 466

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4329> which encodes the amino acid sequence <SEQ ID 4330>. Analysis of this protein sequence reveals the following:

```

50     Possible site: 60
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0275 (Affirmative) < succ>
55     bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

60     Identities = 440/468 (94%), Positives = 456/468 (97%)

      Query: 1  MSSGKIAQVVGPPVVDVVFASGDKLPEINNALLIVYKNGDKSQKVVLEVALELGDGLVRTIA 60

```

-1556-

MSSGKIAQVVGPVVDV+FASGDKLPEINNALIVYK+ DK QK+VLEVALELGDG+VRTIA
 Sbjct: 1 MSSGKIAQVVGPVVDVVFASGDKLPEINNALIVYKSDKKQKIVLEVALELGDGMVRTIA 60

Query: 61 MESTDGLTRGLEVLDTGRAISVPVGKDTLGRVFNVLGDAIDLEEPFAEDAERQPIHKKAP 120
 MESTDGLTRGLEVLDTGRAISVPVGK+TLGRVFNVLG+ IDLEEPFAED +RQPIHKKAP
 Sbjct: 61 MESTDGLTRGLEVLDTGRAISVPVGKETLGRVFNVLGETIDLEEPFAEDVDROPIHKKAP 120

Query: 121 SFDELSTSSIELETGIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIQAQEHGGIS 180
 SFDELSTSSIELETGIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIQAQEHGGIS
 Sbjct: 121 SFDELSTSSIELETGIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIQAQEHGGIS 180

Query: 181 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE 240
 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE
 Sbjct: 181 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE 240

Query: 241 GQDVLLFIDNIFRFTQAGSEVSALLGRMPGSAVGYPQTLATEMGQLQERITSTKKGSVTSI 300
 GQDVLLFIDNIFRFTQAGSEVSALLGRMPGSAVGYPQTLATEMGQLQERITST+KGSVTSI
 Sbjct: 241 GQDVLLFIDNIFRFTQAGSEVSALLGRMPGSAVGYPQTLATEMGQLQERITSTQKGSVTSI 300

Query: 301 QAIYVPADDYTDPAAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRAL+PEIVGDEH 360
 QAIYVPADDYTDPAAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRAL+PEIVG+EH
 Sbjct: 301 QAIYVPADDYTDPAAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRALSPEIVGEEH 360

Query: 361 YEVATEVQVRVLRQYRELQDIIAILGMDLSDEEKTIVGRARRIQFFLSQNFNVAETFTGQ 420
 Y VATEVQVRVLRQYRELQDIIAILGMDLSDEEKTIVGRARRIQFFLSQNFNVAE FTG
 Sbjct: 361 YAVATEVQVRVLRQYRELQDIIAILGMDLSDEEKTIVGRARRIQFFLSQNFNVAEQFTGL 420

Query: 421 PGSYVPVEETVRGFKELDGKHDQIPEDAFRMVGGIEDVIAKAEKMN 468
 PGSYVPV +TVRGFKEL+GK+D++PEDAFR VG IEDVI KAEKM +
 Sbjct: 421 PGSYVPVADTVRGFKELLEGKYDELPEDAFRSVGPIEDVIKAEKMGF 468

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1411

35 A DNA sequence (GBSx1496) was identified in *S.agalactiae* <SEQ ID 4331> which encodes the amino acid sequence <SEQ ID 4332>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1889(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA23754 GB:AB009314 proton-translocating ATPase, gamma subunit
 [Streptococcus bovis]
 Identities = 252/293 (86%), Positives = 278/293 (94%), Gaps = 2/293 (0%)

Query: 1 MAGSLSEIKDKILSTEKTSKITSAMQMVSSAKLVKSEQAARDFQVYASKIRQITINLLKS 60
 MAGSLSEIK KI+ST+KTS IT AMQMV+AKL KSEQAA+DFQVYASKIRQITT+LLKS
 Sbjct: 1 MAGSLSEIKGKIISTQKTSHTGAMQMVSAKLTKSEQAAKDFQVYASKIRQITIDLLKS 60

Query: 61 DLVSGSDNPMMLSSRPVKKTGYIVITSDKGLVGGYNSKILKAMMDTTDYHTENDDYAIIS 120
 +LV+GS NPML++RPVKKTGYIVITSDKGLVGGYNSKILKAMMD I +YH ++ +YAI+
 Sbjct: 61 ELVNGSKNPMMLAARFVKKTGYIVITSDKGLVGGYNSKILKAMMDLIBEYH-QDGNAYIAI 119

Query: 121 IGSVGSDFFKARGMNVSEFELRGLEDQPSFDQVGKIIAQAVEMYKNELFDELYVCYNHHVN 180
 IG +G+DFFKARGMNV FELRGLEDQPSF+QVG IIA++VEMYKNELFDELYVCYNHHVN
 Sbjct: 120 IGGIGADFFKARGMNVVFEFELRGLEDQPSFEQVGNIIAKSVEMYKNELFDELYVCYNHHVN 179

Query: 181 SLTSQVRMQMLPIKELDAEEASEDRVITGFELEPNREVILEQLLPQYTESLIYGAIIDA 240

-1557-

SLTSQVR+QQMLPI ELDA+EA+E+ V +GFELEPNRE+ILEQLLPQYTESLIYGAI+DA
 Sbjct: 180 SLTSQVRVQQMLPIAELDADEAAEEGV-SGFELEPNREMILEQLLPQYTESLIYGAIVDA 238

Query: 241 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRQAAITQEITEIVAGANALE 293
 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRQAAITQEITEIVAGANALE
 Sbjct: 239 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRQAAITQEITEIVAGANALE 291

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4333> which encodes the amino acid sequence <SEQ ID 4334>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1969(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 251/293 (85%), Positives = 275/293 (93%), Gaps = 2/293 (0%)

Query: 1 MAGSLSEIKDKILSTEKTSKITSAMQMVSSAKLVKSEQAARDFQVYASKIRQITTNLLKS 60
 MAGSLSEIK KI+STEKTSKITSAM+MVSSAKLVKSEQAARDFQ+YASKIRQITT+LLKS
 Sbjct: 1 MAGSLSEIKAKIISTEKTSKITSAMRMVSSAKLVKSEQAARDFQIVYASKIRQITTDLLKS 60

Query: 61 DLVSGSDNPMLSSRPVKKTGYIVITSDKGLVGGYNSKILKAMMDTITDYHTENDDYAIIS 120
 +L GSDNPML SRPVKKTGYIVITSDKGLVGGYNSKILK++MD IT+YH + DY IIS
 Sbjct: 61 ELTIGSDNPMLVSRPVKKTGYIVITSDKGLVGGYNSKILKSVMDMITEYHADG-DYEIIS 119

Query: 121 IGSVGSDFFKARGMNVFELRGLDQPSFDQVGKIIAQAVEMYKNELFDELYVCYNHHVN 180
 IGSVGSDFFKARGMNV+FELRGL DQPSF+QV +II+Q+V+M+ NE+FDELYVCYNHHVN
 Sbjct: 120 IGSVGSDFFKARGMNVAFELRGLADQPSFEQVRQIISQSVDMFVNEIFDELYVCYNHHVN 179

Query: 181 SLTSQVRMQQMLPIKELDAEEASEDRVITGFELEPNREVILEQLLPQYTESLIYGAIIDA 240
 SLTSQVR+QQMLPI +L A+EA+E+ V TGFLEPNR IL+QLLPQ+TESLIYGAIIDA
 Sbjct: 180 SLTSQVRVQQMLPISDLVADEAAEEGV-TGFLEPNRHDILDQLLPQFTESLIYGAIIDA 238

Query: 241 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRQAAITQEITEIVAGANALE 293
 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRQAAITQEITEIVAGANALE
 Sbjct: 239 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRQAAITQEITEIVAGANALE 291

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1412

A DNA sequence (GBSx1497) was identified in *S.agalactiae* <SEQ ID 4335> which encodes the amino acid sequence <SEQ ID 4336>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1963(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1558-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1413

A DNA sequence (GBSx1498) was identified in *S.agalactiae* <SEQ ID 4337> which encodes the amino acid sequence <SEQ ID 4338>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3146(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to the alpha subunit of the proton-translocating ATPase from *S.bovis*:

15 >GP:BAA23753 GB:AB009314 proton-translocating ATPase, alpha subunit
 [Streptococcus bovis] Length = 501
 Identities = 482/501 (96%), Positives = 497/501 (98%)

20 Query: 1 MAINAQEISALIKKQIEDFQPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEFSNGAY 60
 MAINAQEISALIKKQIE+FQPNFDVTETG+VTYIGDGIARARGLDNAMSGELLEFSNGA+
 Sbjct: 1 MAINAQEISALIKKQIENFQPNFDVTETGVVTTYIGDGIARARGLDNAMSGELLEFSNGAF 60

25 Query: 61 GMAQNLESNDVGIIILGDFSEIREGDVVKRTGKIMEVPVGEAMIGRVVNPLGQPVDGLGE 120
 GMAQNLESNDVGIIILGDFS IREGD VKRTGKIMEVPVGEA+IGRVVNPLGQPVDGLG+
 Sbjct: 61 GMAQNLESNDVGIIILGDFSTIREGDEVKRTGKIMEVPVGEALIGRVVNPLGQPVDGLGD 120

30 Query: 121 IETTATRPVETPAPGVMQRKSVFEPLQTLKALDALVPIGRGQRELIIGDRQTGKTSVAI 180
 I+TTATRPVETPAPGVMQRKSV EPLQTLKALDALVPIGRGQRELIIGDRQTGKTSVAI
 Sbjct: 121 IKTTATRPVETPAPGVMQRKSVSEPLQTLKALDALVPIGRGQRELIIGDRQTGKTSVAI 180

35 Query: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRYGALDYTTIVVTASASQPSPLLFIAPY 240
 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRYGALDYTTIVVTASASQPSPLL+IAPY
 Sbjct: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRYGALDYTTIVVTASASQPSPLLYIAPY 240

40 Query: 241 AGVAMAEFMYNGKHLIVYDDLKQAVAYRELSLLLRPPGREAYPGDVFYLSRLLER 300
 AGVAMAEFMYNGKHLIVYDDLKQAVAYRELSLLLRPPGREAYPGDVFYLSRLLER
 Sbjct: 241 AGVAMAEFMYNGKHLIVYDDLKQAVAYRELSLLLRPPGREAYPGDVFYLSRLLER 300

 Query: 301 SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG 360
 SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG
 Sbjct: 301 SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG 360

 Query: 361 SSVSRVGGAAQIKAMKRVAGTLRLDLASYRELEAFTQFGSDLLDAATQAKLNRGRRTVEVL 420

-1559-

bacterial cytoplasm --- Certainty=0.3654(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 477/501 (95%), Positives = 490/501 (97%)

```

Query: 1  MAINAQEISALIKKQIEDFQPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEFSNGAY 60
+AINAQEISALIKKQIE+QPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEFSNGAY
10 Sbjct: 1  LAINAQEISALIKKQIENFQPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEFSNGAY 60

Query: 61  GMAQNLESNDVGIIILGDFSEIREGDDVVKRTGKIMEVPVGEAMIGRVVNPLGQPDGLGE 120
GMAQNLESNDVGIIILGDFS IREGDVKRTGKIMEVPVGEA+IGRVVNPLGQPDGLGE+
15 Sbjct: 61  GMAQNLESNDVGIIILGDFSAIREGDDVVKRTGKIMEVPVGEALIGRVVNPLGQPDGLGE 120

Query: 121 IETTATRPVETPAPGVMQRKSVFEPLQTGLKALDALVPIGRGQRELIIGDRQTGKTSVAI 180
IETT RPVETPAPGVMQRKSV EPLQTGLKALDALVPIGRGQRELIIGDRQTGKTSVAI
Sbjct: 121 IETTGFPRPVETPAPGVMQRKSVSEPLQTGLKALDALVPIGRGQRELIIGDRQTGKTSVAI 180

Query: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRKYGALDYTIVVTASASQPSPLLFIAPY 240
DAILNQKGQDMICIYVAIGQKESTVRTQVETLR+YGALDYTIVVTASASQPSPLLFIAPY
20 Sbjct: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRKYGALDYTIVVTASASQPSPLLFIAPY 240

Query: 241 AGVAMAEFMYGKHVLIYVDDLSKQAVAYRELSLLRRPPGREAYPGDVFYLHSRLLER 300
AGVAMAEFMY GKHVLIYVDDLSKQAVAYRELSLLRRPPGREAYPGDVFYLHSRLLER
25 Sbjct: 241 AGVAMAEFMYGKHVLIYVDDLSKQAVAYRELSLLRRPPGREAYPGDVFYLHSRLLER 300

Query: 301 SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDGGQIFLQENLFNSGIRPAIDAG 360
SAKVSD LGGGSITALPFIETQAGDISAYIATNVISITDGGQIFLQENLFNSGIRPAIDAG
30 Sbjct: 301 SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDGGQIFLQENLFNSGIRPAIDAG 360

Query: 361 SSVSRVGGAAQIKAMKRVAGTFLRLDLASYRELEAFTQFGSDLLDAATQAKLNRGRRTVEVL 420
SSVSRVGG+AQIKAMK+VAGTFLRLDLASYRELEAFTQFGSDLLDAATQAKLNRGRRTVE+L
35 Sbjct: 361 SSVSRVGGSAQIKAMKRVAGTFLRLDLASYRELEAFTQFGSDLLDAATQAKLNRGRRTVEIL 420

Query: 421 KQPLHKPLPVEKQVVILYALTHGFLDDVPVDILAFEEALYDYFDAHYNDLFETIRTTKD 480
KQPLHKPLPVEKQVVILYALTHGFLDDVPV+DILAFEEALYDYFD HY++LFETIRTTKD
Sbjct: 421 KQPLHKPLPVEKQVVILYALTHGFLDDVPVDILAFEEALYDYFDVHYNDLFETIRTTKD 480

Query: 481 LPPEEAELDAAIQAFKDSQPK 501
LPPEA LDAAI+AFK+ S FK
40 Sbjct: 481 LPPEEAALDAAIKAFKEHSNFK 501

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 45 vaccines or diagnostics.

Example 1414

A DNA sequence (GBSx1499) was identified in *S.agalactiae* <SEQ ID 4341> which encodes the amino acid sequence <SEQ ID 4342>. Analysis of this protein sequence reveals the following:

Possible site: 55
 50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1896(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA23752 GB:AB009314 proton-translocating ATPase, delta subunit
 [Streptococcus bovis]
 60 Identities = 98/178 (55%), Positives = 127/178 (71%)

-1560-

Query: 1 MNKKTQALIEQYSKSLVEVAIEHKIVEKIQQEVAAALIDIFETSELEGLVSSLAVSHDEKQ 60
 M+KKTQAL+EQY+KSLVE+AIE ++Q E AL+ +FE + L LSSL VS DEK
 Sbjct: 1 MDKKTQALVEQYAKSLVEIAIEKDSLAELOSETALLSVFEETNLADFLSSLVVSDEKV 60

5 Query: 61 HFVKTLQTSCSTYLVNFLEVIVQNEREALLYPILKSVDQELIKVNGQYPIQITTAVALS 120
 V+ LQ S S Y+ NFLEVI+QNEREA L IL+ V ++ + Q+ I +TTAVAL+
 Sbjct: 61 KLVRLLOESSVYMNNFLEVILQNEREAFLKAILEGVQKDFVIATNQHDIVVTITAVALT 120

10 Query: 121 EQKERLFDIAKTKLALPNGQLVEHIDPSIVGGFVVNANNKVIDASVRNQLHQFKMKLK 178
 EQKER+ + K + G+LVE+ID SI+GGFV+N NNKVID S+R QL +FKM LK
 Sbjct: 121 EQKERILALVAEKFGVKAGKLVENIDESILGGFVINVNKVIDTSIRRLQOEFKMNLK 178

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4343> which encodes the amino acid sequence <SEQ ID 4344>. Analysis of this protein sequence reveals the following:

15 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1668(Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/178 (48%), Positives = 125/178 (69%)

25 Query: 1 MNKKTQALIEQYSKSLVEVAIEHKIVEKIQQEVAAALIDIFETSELEGLVSSLAVSHDEKQ 60
 M KK QALIEQY+KSLVEVA EH ++ +Q +V A+++ F T+ L+ LSS AV H EK
 Sbjct: 1 MTKKEQALIEQYAKSLVEVASEHSLDALQADVLAILLETFTVTTNLDQSLSSQAVPHAKEI 60

30 Query: 61 HFVKTLQTSCSTYLVNFLEVIVQNEREALLYPILKSVDQELIKVNGQYPIQITTAVALS 120
 + L+ + S Y+ NFL +I+QNEREA LY +L++V E+ V+ QY + +T+++ L+
 Sbjct: 61 KLLTLLKGNNNSVYMNFLNLILQNEREAYLYQMLQAVLNEIAIVSNQYDVTVTSSSLPLTE 120

35 Query: 121 EQKERLFDIAKTKLALPNGQLVEHIDPSIVGGFVVNANNKVIDASVRNQLHQFKMKLK 178
 EQK R+ + K A+ G+L+E +DPS++GGF+++ NNKVID S+R QL FKM LK
 Sbjct: 121 EQKSRVRAVVAKKFAVTAGRLIEKVDPSLIGGFIIISVNNKVIDTSIRRLQAFKMNK 178

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1415

A DNA sequence (GBSx1500) was identified in *S.agalactiae* <SEQ ID 4345> which encodes the amino acid sequence <SEQ ID 4346>. This protein is predicted to be ATP synthase b chain (atpF). Analysis of this protein sequence reveals the following:

45 Possible site: 33
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD13379 GB:U31170 ATPase, b subunit [Streptococcus mutans]
 Identities = 103/165 (62%), Positives = 130/165 (78%)

55 Query: 1 MSILINSTIGDIIIVSGSVLLLFIILIKTFAWKQITGIFEAREQKIANDIDTAEQARQQA 60
 MS LIN T++G+++IV+GS +LL +L+K FAW Q+ IF+ RE+KIA DID AE +RQ A
 Sbjct: 1 MSTLINGTSLGNLLIVTGSFILLLLLVLKFAWSQLAAIFKTREEKIAKDIDDAENSRQNA 60

-1561-

Query: 61 EAFATKREEELSNKTEANQIIDNAKETGLAKGDIIEAKTEADRLKEKAHQDIAQNKA 120
 + KR+ EL+ AK EA QIIDNAKETG A+ +II+EA EA RLK+KA+QDIA +KA
 Sbjct: 61 QVLENKROVELNQAKDEAAQIIDNAKETGKAQESKIITEAHEEAGRLKDKANQDIATSKA 120

Query: 121 EALADVKGEVADLTVLLAEKIMVSNLDKEAQSNIIDSYIKKLGA 165
 EAL+ VK +VADL+VLLAEKIM NLDK AQ +LIDSY+ KLGA
 Sbjct: 121 EALSSVKADVADLSVLLAEKIMAKNLDKTAQGDILIDSYLDKLGA 165

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4347> which encodes the amino acid sequence <SEQ ID 4348>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have a cleavable N-term signal seq.

- 15 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the databases:

>GP:AAD13379 GB:U31170 ATPase, b subunit [Streptococcus mutans]
 Identities = 88/159 (55%), Positives = 122/159 (76%)

- 25 Query: 6 GELVGNFILVTGSIIVLLLLIKKFAWGAIESILQTRSQQISRDIQAEQSRLSAQQLEAK 65
 G +GN ++VTGS I+LLLL+KKFAW + +I +TR ++I++DID AE SR +AQ LE K
 Sbjct: 7 GTSLGNLLIVTGSPILLLLLVKKFAWSQLAAIFKTREEKIAKDIDDAENSRQNAQVLENK 66
- 30 Query: 66 SQANLDASRLQASKIISDAKEIGQLQGDKLVAEATDEAKRLKEKALTDIEQSKSDAISAV 125
 Q L+ ++ +A++II +AKE G+ Q K++ EA +EA RLK+KA DI SK++A+S+V
 Sbjct: 67 RQVELNQAKDEAAQIIDNAKETGKAQESKIITEAHEEAGRLKDKANQDIATSKAEALSSV 126
- 35 Query: 126 KTEMSDLTVLLAEKIMGANLDKTAQSQLIDSYLDLGEA 164
 K +++DL+VLLAEKIM NLDKTAQ LIDSYLD LG+A
 Sbjct: 127 KADVADLSVLLAEKIMAKNLDKTAQGDILIDSYLDKLGA 165

An alignment of the GAS and GBS proteins is shown below.

Identities = 81/156 (51%), Positives = 115/156 (72%)

- 40 Query: 10 IGDIIIVSGSVLLLFILIKTFWAKQITGIFEAREQKIANDIDTAEQARQQAFAFATKREE 69
 +G+ I+V+GSV++L +LIK FAW I I + R Q+I+ DID AEQ+R A+ K +
 Sbjct: 9 VGNFILVTGSIIVLLLLIKKFAWGAIESILQTRSQQISRDIQAEQSRLSAQQLEAKSQA 68
- 45 Query: 70 ELSNAKTEANQIIDNAKETGLAKGDIIEAKTEADRLKEKAHQDIAQNKAELADVKGE 129
 L ++ +A++II +AKE G +GD++++EA EA RLKEKA DI Q+K++A++ VK E
 Sbjct: 69 NLDASRLQASKIISDAKEIGQLQGDKLVAEATDEAKRLKEKALTDIEQSKSDAISAVKTE 128
- 50 Query: 130 VADLTVLLAEKIMVSNLDKEAQSNIIDSYIKKLGA 165
 ++DLTVLLAEKIM +NLDK AQS LIDSY+ LG+A
 Sbjct: 129 MSDLTVLLAEKIMGANLDKTAQSQLIDSYLDLGEA 164

SEQ ID 4346 (GBS169) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 6; MW 18kDa).

- The GBS169-His fusion product was purified (Figure 200, lane 11) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 250). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1562-

Example 1416

A DNA sequence (GBSx1501) was identified in *S.agalactiae* <SEQ ID 4349> which encodes the amino acid sequence <SEQ ID 4350>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -11.73    Transmembrane    20 - 36 ( 14 - 42)
    INTEGRAL    Likelihood = -5.20     Transmembrane    207 - 223 ( 206 - 228)
    INTEGRAL    Likelihood = -4.35     Transmembrane    78 - 94 ( 73 - 97)
10  INTEGRAL    Likelihood = -4.09     Transmembrane    113 - 129 ( 113 - 133)
    INTEGRAL    Likelihood = -2.39     Transmembrane    174 - 190 ( 174 - 190)

----- Final Results -----
    bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA23750 GB:AB009314 proton-translocating ATPase, a subunit
    [Streptococcus bovis]
20  Identities = 149/238 (62%), Positives = 180/238 (75%)

Query: 1  MESTSNPTVSFLGIDFDLTILAMSLLTITITIFILVFWASRKMTIKPKGQNVLEYVYELV 60
        ME++ NPT  GI+FDLTILAMSLLT+ I F ++FWA+RKMT+KPKGQON +EYVYE V
Sbjct: 1  METSVNPTAHVFGIEFDLTILAMSLLTVIIISFGIIFWATRKM TLKPKGQNFIEYVYEFV 60

25  Query: 61  NNTISQNLGHYTKNYSLLMFILFSFVFIANNLGLMTSLKTHEHNFWTSPTANFGVDITLS 120
        NTI  NLG YT  YSLLMF  F F+ IANNLGL+  L++ ++NFWTSPT+  VD T S
Sbjct: 61  QNTIKPNLGEYTPKYSLLMFTFFFFILIANNLGLLVKLESEDYNFWTSPTSTIMVDCIWS 120

30  Query: 121 LLVAFICHIEGIRKKGIGGYLKGFLSPTPAMLPNNLLEEVTVNASLALRLFGNIFSGEVV 180
        L+VA + H+EG+RKKG+  YLKG+LSP P MLPNN+LE+ TNV SLALRLFGNI++GEVV
Sbjct: 121 LLVAIVVHVHVGVRKKGVKAYLKGYSPPFMMMLPMNILEQFTNVLSLALRLFGNIYAGEVV 180

35  Query: 181 TGLLLQLAVLSPFTGPLAFALNIVWTAFSMFIGFIQAYVFIISSSYIGHKVHGDEEE 238
        T L++  S  P A ALN+ W AFS FIG IQAYVF ILSS YI  K+  DE+E
Sbjct: 181 TALIVGFGTKSLIFAPFALALNLAWVAFSAFIGCIQAYVFTILSSKYISEKLPEDEDE 238

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4351> which encodes the amino acid sequence <SEQ ID 4352>. Analysis of this protein sequence reveals the following:

```

40  Possible site: 33
    >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -4.73     Transmembrane    79 - 95 ( 72 - 97)
    INTEGRAL    Likelihood = -4.35     Transmembrane    115 - 131 ( 112 - 132)
45  INTEGRAL    Likelihood = -2.13     Transmembrane    200 - 216 ( 197 - 216)

----- Final Results -----
    bacterial membrane --- Certainty=0.2890(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 124/239 (51%), Positives = 169/239 (69%), Gaps = 3/239 (1%)

55  Query: 1  MESTSNPTVSFLGIDFDLTILAMSLLTITITIFILVFWASRKMTIKPKGQNVLEYVYELV 60
        ME  P +  I F+LT+LA+ ++TI I+F VFWASR+M +KP+GKQ LEY+  V
Sbjct: 1  MEEAKIPMLKLGPIITFNLTLLAVCIVTIAIVFAFVFWASRQMKLKPEGKQTALEYLISFV 60

    Query: 61  NNTISQNLGH-YTKNYSLLMFILFSFVFIANNLGLMTSLKT-HEHNFWTSPTANFGVDIT 118
        +  ++L H  K+YSL+L+F +F FV +ANNLGL T L+T + +N WTSPTAN  D+
60  Sbjct: 61  DGIGEBHLDHNLQKSYSLLLFTIFLFAVANNLGLFTKLETVNGYNLWTSPTANLAFDLA 120

```

-1563-

Query: 119 LSLVAFICHIEGIRKKGIGGYLKGFLSPTPAMLPNNLLEEVNTNASLALRLFGNIFSGE 178
 LSL + + HIEG+R++G+ +LK +P P M PMNLEE TN SLA+RLFGNIF+GE
 Sbjct: 121 LSLFITLMVHIEGVRRRLVAHLKRLATPWP-MTPMNLLEEFNLSLALRLFGNIFAGE 179

5 Query: 179 VVTGLLLQLAVLSPFTGPLAFALNIVWTAFSMFIGFIQAYVFIISSSYIGHKVGDEE 237
 VVTGL++QLA + P+AF +N+ WTAFS+FI IQA+VF L+++Y+G KV+ EE
 Sbjct: 180 VVTGLIVQLANYRVYWPIAFVLNMAWTAFSVFISCIQAFVFTKLTATYLGKKVNESEE 238

A related GBS gene <SEQ ID 8803> and protein <SEQ ID 8804> were also identified. Analysis of this
 10 protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 1
 McG: Discrim Score: -3.50
 GvH: Signal Score (-7.5): -3.36
 Possible site: 29

15 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 5 value: -11.73 threshold: 0.0
 INTEGRAL Likelihood = -11.73 Transmembrane 20 - 36 (14 - 42)
 INTEGRAL Likelihood = -5.20 Transmembrane 207 - 223 (206 - 228)
 20 INTEGRAL Likelihood = -4.35 Transmembrane 78 - 94 (73 - 97)
 INTEGRAL Likelihood = -4.09 Transmembrane 113 - 129 (113 - 133)
 INTEGRAL Likelihood = -2.39 Transmembrane 174 - 190 (174 - 190)
 PERIPHERAL Likelihood = 5.30 156
 modified ALOM score: 2.85

25 *** Reasoning Step: 3

----- Final Results -----

30 bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF01818(301 - 1014 of 1314)
 GP|2662321|dbj|BAA23750.1||AB009314(1 - 238 of 239) proton-translocating ATPase, a subunit
 35 {Streptococcus bovis}
 %Match = 35.0
 %Identity = 62.2 %Similarity = 78.6
 Matches = 148 Mismatches = 51 Conservative Sub.s = 39

40 204 234 264 294 324 354 384 414
 XANCQTLMLPGVGFIERFYFLRSICVYILSKIDDNLEKKEG*GLESTSNPTVSFLGIDFDLTILAMSLTTITIIIFILVFWA
 :|:: ||| :|::| ||||| ||||| :| :|::|
 METSVNPTAHVFGIEFDLTILAMSLTTVIISFGIIFWA
 10 20 30

45 444 474 504 534 564 594 624 654
 SRKMTIKPKGKQNVLEYVYELVNNTISQNLGHYTKNYSLLMFI LFSFVFIANNLGLMTSLKTHEHNFWTSPTANFGVDIT
 :|||:| ||||| :|||:| ||| ||| ||| :| :|::| ||||| :|:: :| ||||| :| |||
 50 TRKMTLKPKGKQNFIEYVYEFVQNTIKPNLGEYTPKYSLLMFTFFFFILIANNLGLLVKLESEDYNFWTSPTSTIMVDCT
 50 60 70 80 90 100 110

55 684 714 744 774 804 834 864 894
 LSLVAFICHIEGIRKKGIGGYLKGFLSPTPAMLPNNLLEEVNTNASLALRLFGNIFSGEVVTGLLLQLAVLSPFTGPLA
 ||::| :|::| ||||| :|||:| ||| ||||| :|::| ||||| ||||| :|::| :| :|
 WSLIVAIVVHVEGVRRKGKGYLKGFLSFPFPMMLPNNILEQFTNVLSLALRLFGNIYAGEVVTALIVGFGTKSLIFAPFA
 130 140 150 160 170 180 190

60 924 954 984 1014 1044 1074 1104 1134
 FALNIVWTAFSMFIGFIQAYVFIISSSYIGHKVGDEEE*EKRGEICQYLLIVQRLVISLSYLALCFSYLS*LRL LHGN
 :|||:| ||| ||| ||||| ||||| ||| :| :|::|
 LALNLAWVAFSAFIGICIQAYVFTILSSKYISEKLPEDEDET
 210 220 230

-1564-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1417

A DNA sequence (GBSx1502) was identified in *Sagalactiae* <SEQ ID 4353> which encodes the amino acid sequence <SEQ ID 4354>. This protein is predicted to be ATP synthase c subunit (atpE). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -4.62    Transmembrane    48 - 64 ( 42 - 65)

----- Final Results -----
          bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA23749 GB:AB009314 proton-translocating ATPase, c subunit [Streptococcus bovis]
Identities = 56/65 (86%), Positives = 59/65 (90%)
```

```
Query: 1  MNLAIALGFAVMGVSIGEGILVANIAKSAARQPEMFSKLQTLMTGVAFIEGTFVFLFA 60
          +NL ILALG AV+GVS+GEGILVANIAKSAARQPEMFSKLQTLMF GVAFIEGTFVFL A
Sbjct: 2  LNLKILALGLAVLGVSLGEGILVANIAKSAARQPEMFSKLQTLMTFLGVAFIEGTFVFLA 61

Query: 61  FTFLV 65
          TF V
Sbjct: 62  STFFV 66
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4355> which encodes the amino acid sequence <SEQ ID 4356>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -5.26    Transmembrane    47 - 63 ( 41 - 64)

----- Final Results -----
          bacterial membrane --- Certainty=0.3102(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD00920 GB:AF001955 UncE [Streptococcus sanguinis]
Identities = 50/66 (75%), Positives = 58/66 (87%), Gaps = 1/66 (1%)

Query: 1  MNPIF-ALALACFGVSLAEGFLMANLFKAASRQPEIIGQLRSLMILGVAFIEGTFVTLV 59
          MN F L ACFGVS+AEG +M+NLFKAASRQPEIIGQLRSL+ILG+AF+EGTFVTL
Sbjct: 1  MNLTFLGLCFACFGVSLAEGFLMANLFKAASRQPEIIGQLRSLILGIAFVEGTFVTLA 60

Query: 60  MAFILK 65
          MAF++K
Sbjct: 61  MAFVIK 66
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 33/62 (53%), Positives = 45/62 (72%)

Query: 5  ILALGFAVMGVSIGEGILVANIAKSAARQPEMFSKLQTLMTGVAFIEGTFVFLFAFTFLVR 66
          I AL A GVS+ EG L+AN+ K+A+RQPE+ +L++LM GVAFIEGTFV F+++
Sbjct: 4  IFALALACFGVSLAEGFLMANLFKAASRQPEIIGQLRSLMILGVAFIEGTFVTLVMAFILK 65
```

-1565-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1418

A DNA sequence (GBSx1503) was identified in *S.agalactiae* <SEQ ID 4357> which encodes the amino acid sequence <SEQ ID 4358>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.2562(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1419

20 A DNA sequence (GBSx1504) was identified in *S.agalactiae* <SEQ ID 4359> which encodes the amino acid sequence <SEQ ID 4360>. This protein is predicted to be bacterial glycogen synthase (glgA). Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1574(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA19591 GB:D87026 bacterial glycogen synthase [Bacillus
 stearothermophilus]
Identities = 220/475 (46%), Positives = 312/475 (65%), Gaps = 1/475 (0%)

35 Query: 1 MKIMFVAEAGAPFAKT'GGLGDVIGALPKSLSKKHGDVAVVMPYYDMVDQKFGDQIENLMY 60
 MK++F +E APFAK+GGL DV GALPK L + G D V++P Y+ + ++ +++ +
Sbjct: 1 MKVLFVAVSECAPFAKSGGLADVAGALPKELRRLGIDARVMLPKYETIAPWKKMKKVAE 60

40 Query: 61 FYTDVGWRHQYVGKRLSQDNVTFYFIDNQYYFYRGHVYGDWDDGERFAYFQLAALELME 120
 VGWR QY GV+ L D V +YFIDN+YYF R +YG +DDGERFAYF A LE++
Sbjct: 61 LIVPVGWRQYCGVEELRHDGVIIYFIDNEYFYFKRPQLYGHYDDGERFAYFCRAVLEVL 120

45 Query: 121 KIDFIPDVLHVHDYHTAMIPFLLKEKYHWIQAYNNIRAVFTIHNIEFQGFQGPMLGDLF 180
 +I F PDV+H HD+HT M+PFLL+E+Y Y ++R VFTIHN++FQG F +L DL
Sbjct: 121 BIQFQPDVIHCHDWHGTGMVPFLLREQYRHELFYVDMRTVFTIHNLFQGLFPRGILEDLL 180

50 Query: 181 GVGAERYEDGTLRWNNCLNWMKAAILYSRVTTVSPSYANEIKTPEFGKGLDQIMRMEAG 240
 + + L + C+++MK A++ SD +TTVSP+Y EI+T +G+ LD ++R
Sbjct: 181 NLDGRYFTVDHLEFYGCVSFMKGALVASDLITTVSPTYKEEIQATAYYGERLDGLLRARRD 240

Query: 241 KLSGIVNGIDSDLLNPETDAFLPYHFSKSNLEGGKIKNKLALQENLGLPQDKNVPLIGIVS 300
 L GI+NGID + NPE D FL +S E K NK ALQ GLP+ +VPLI +V+
Sbjct: 241 DLLGILNGIDDEFYNPEADPFLTATYSVHTRERKQLNKRALQRQFGLPEWDDVPLIAMVT 300

-1566-

Query: 301 RLTDQKGFDIASELDNMLQQDIQMVILGTGYHHFEETFSYFASRYPEKLSANITFDLRL 360
 R+T QKG D++ M+ +D+Q+V+LGTG FE+ FS A+ YP K+ I F L
 Sbjet: 301 RMTAQKGLDLVTCVFHEMSEDMLVVLGTGDWRFEQFFSQMAAAYPGKVGVIYGFHEPL 360

5 Query: 361 AQQIYAASDIFMMPSAFEPCGLSQMMAMRYGSLPLVHEVGGLKDTTVAFNQFDGSGTGFS 420
 A QIYA +D+F++PS FEPCGLSQM+A+RYG++P+V E GGL DTV ++N+ G GFS
 Sbjet: 361 AHQIYAGADLFLIPSLFEPCGLSQMIALRYGTIPIVRETGGLNDTVQSYNEITKEGNGFS 420

10 Query: 421 FNHFSGYWLMQTLKLALLEVYNDYPEAWKKLQWQAMSKDFSWDTACVAYEQLYQQL 475
 F +F+ + ++ T++ AL Y P W++L +AM D+SW + Y+Q Y+QL
 Sbjet: 421 FTNFNAHDMLYTIRRALSFYRQ-PSVWEQLTERAMRGDYSWRRSANQYKQAYEQL 474

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1420

A DNA sequence (GBSx1505) was identified in *S.agalactiae* <SEQ ID 4361> which encodes the amino acid sequence <SEQ ID 4362>. This protein is predicted to be a subunit of ADP-glucose pyrophosphorylase. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3492 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA19590 GB:D87026 subunit of ADP-glucose pyrophosphorylase
 [Bacillus stearothermophilus]
 Identities = 59/178 (33%), Positives = 111/178 (62%), Gaps = 1/178 (0%)

Query: 37 SAEIYVIDTPWLEIKMEEEAQNNPEPRKLRFLRLDLIVESNALAFETYTGYLENISSIKSY 96
 S E+Y+++T L++ + + +N+ + ++RD + +EY+GY + I S++ Y+
 Sbjet: 157 SLEMYLLETSLLLDLIADY-KNHGYYSIVDVIRYHRSLSICEYEYSGYAAVIDSVEQYF 215

Query: 97 DANMDMLTPNKFYSLFFSNQKVYTKVKNERATYFDKQSNVSNSQLASGSIKGYLDHSIV 156
 ++M++L + + LF + +YTKVK+E T + ++ NV S +A+G +I+G +++S++
 Sbjet: 216 RSSMELLDRDVWEQLFLPSHPITVKVDEPPTKYGREGNVKRSMIANGCVIEGTVENSVL 275

Query: 157 SRNCLLEKGRVNSIIFPKVKIGEGATIENTIIDKCVKVASGVTLKGSGLDKPLVIPK 214
 R+ + KG V NSII K +IG+G ++ IIDK KV GV LKG+ ++P ++ K
 Sbjet: 276 FRSVKIGKGAVERNSIIMQKCQIGDGCVLGDGVIIDKDAKVERGVVLKGTKEQPFIVRK 333

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1421

A DNA sequence (GBSx1506) was identified in *S.agalactiae* <SEQ ID 4363> which encodes the amino acid sequence <SEQ ID 4364>. This protein is predicted to be subunit of ADP-glucose pyrophosphorylase (glgC-1). Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq

-1567-

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9765> which encodes amino acid sequence <SEQ ID 9766> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:BAA19589 GB:D87026 subunit of ADP-glucose pyrophosphorylase
 [Bacillus stearothermophilus]
 Identities = 195/352 (55%), Positives = 259/352 (73%)

15 Query: 7 MKNEMLALILAGGQGTIRLQSLTQSIAPAVQFGGRYRIIDFALSNCANSGINNVGVITQY 66
 MK + +A++LAGGQG+RL LT +IAKPAV FCG+YRIIDF LSNCSNGI+ VGV+TQY
 Sbjct: 1 MKKKCIAMLLAGGQGSRLRLTNNIAKPAVPFGGKYRIIDFTLSNCTNSGIDTVGVLTQY 60

20 Query: 67 QPLELNTHIGNSSWGLDGIDSGVTVLQPYSTATENRWFQGTSHAIYQNIIDYIDRINPEY 126
 QPL L+++IG GS+W LD + GVTVL PYS + G +W++GT++A+YQNI+YI++ NP+Y
 Sbjct: 61 QPLLLHSYIGISAWDLDRRNGGVTVLPPYSVSSGVKWEYGTANAVYQNIYIEQYNPDY 120

25 Query: 127 VLILSGDHIYKMNYYDMLQTHKDNLASLTVAVLDPVLPKEASRFGIMNTDSNDRIVEFEK 186
 VL+LSGDHIYKM+Y ML H A +T++V++VP +EASRFGIMNT+ IVEF EK
 Sbjct: 121 VLVLSGDHIYKMDYQHMLDYHIAKQADVTISVIEVPWEERFGIMNTNEEMEIVEFAEK 180

30 Query: 187 PEHPKSTKASMGIIYFDWKRLRTVLIDGKNGIDMSDFGKNVIPAYLESGERVYTYNFDG 246
 P PKS ASMGIIYF+W L+ L N DFGK+VIP L +R + Y F+G
 Sbjct: 181 PARPKSNLASMGIIYFNPWLLKQYLQIDNANPHSSHDFGKDVIPMLLEKKRPFAYPFEG 240

35 Query: 247 YWKDVGTIESLWEANMEYIGEDNKLHSRDRSWKIYSKNIAPPNFMTEANVKDSLVDG 306
 YWKDVGT++SLWEANM+ + E+N+L DRSW+IYS N PP +++ +A V DSLV +G
 Sbjct: 241 YWKDVGTVKSLEWANDLLDENNELDLFDRSWRIYSVNPQPPQYISPEAEVSDSLVNEG 300

Query: 307 CFVAGNVEHSILSTNVQVKPNAIKDSFVMSGATIGEGAKINRAIIGEDAVI 358
 C V G VE S+L V++ A++K+S +M GA + EGA + RAI+ D++I
 Sbjct: 301 CVVEGTVERSFLFQGVRIKGAVVKESVIMPGAASEGAYVERAIVTPDSII 352

There is also homology to SEQ ID 2660.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1422

A DNA sequence (GBSx1507) was identified in *S.agalactiae* <SEQ ID 4365> which encodes the amino acid sequence <SEQ ID 4366>. Analysis of this protein sequence reveals the following:

45 Possible site: 37
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.2844(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:CAA78440 GB:Z14057 1,4-alpha-glucan branching enzyme [Bacillus
 caldolyticus]
 Identities = 272/616 (44%), Positives = 371/616 (60%), Gaps = 14/616 (2%)

Query: 6 ELYTFGIGENFHLQNYLGVSSENGSFC----FRVWAPNAENVQVIGDFTDWRNRPLQMNK 61
 E+Y F G + G H G F VWAP+A V+++G F DW + K

-1568-

Sbjct: 10 EVYLFHEGRLYQSYELFGAHVIRGGGAVGTRFCVWAPHAREVRLVGSFNDWNGTNSPLTK 69

Query: 62 -NQAGVWEANSLDAREGDLYKYLVTTRKGGQVVEKIDPMAYMERRPGTASVIKVLNRNKKW 120
N GWV + EG LYKY + G+V+ K DP A Y E RP TAS++ L+ +W

5 Sbjct: 70 VNDEGVWTIVVPENLEGHLYKYEIITPDGRVLLKADPYAFYSELRPHTASTVYDLKGYEW 129

Query: 121 EDGLWMGRKRRLGFGQKRPINIYEVHAGSWKKDDFGHPMTFSQLKDYLIPLYVEMNYTHVE 180
D W +++R +P+ IYE+H GSWKK G T+ ++ D LIPY++E +TH+Æ

10 Sbjct: 130 NDSWQQRKKRRRIYDQPMVIYELHFGSWKKKPDGRFVYTYREMADELIPYVLERGFTHIE 189

Query: 181 FMPLMAHPLDMSWGYQLMGYFAFHTYGTPEEFQDFVEACHKNNIGVLVDWVPGHFIQND 240
+PL+ HPLD SWGYQ GY++ YGTP +F FV+ CH+ +GV++DWVPGHF ++

Sbjct: 190 LLLPLVEHPLDRSWGYQGTGYYSVTSRYGTPHDFMYFVDRCHQAGLGVIIDWVPGHFCKDA 249

15 Query: 241 DALAYFDGTATYEQNHDRAHNVYRWGALNFDLGKNQVQSFLISSALFWIEHYHIDGIRVD 300
L FDG TYEY N NY WG NFDLGK +V+SPLIS+ALFW+E+YH+DG RVD

Sbjct: 250 HGLYMFDCGAPTYRYANEKDRENVWGTANFDLGKPEVRSFLISNALFWLEYHYHVDGFRVD 309

20 Query: 301 AVSNMMLYLDYDEGPWEANQFGDNRLNLEGYHFLRKLNVKIKERHPNVMMIAEESTASTPIT 360
AV+NMLY ++ +E N FLR+LN+ + PNV MIAE+ST +T

Sbjct: 310 AVANMMLYWFNDRLYE-----NPYAVEFLRQLNEAVFAYDPNVMMIAEDSTDWPRVT 361

25 Query: 361 KDLESGLGLGDFKWNMGWMDILRFYEDPLYRQYDFNLVTFSEFMYIFNENFVLAESHDE 420
GGLGF++KWNMGWMD+L++ E P R+Y N V+FS +Y ++ENF+L FSHDE

Sbjct: 362 APTYDGGGLGFNYKWNMGWMDMLKYMETPPHERKYAHNQVSFSLLYAYSENFLEFESHDE 421

30 Query: 421 VVHGKKSMMHKMGDRYNQFAGLRNLYAYQMCHPGKLLFMGSEFGQFLEWKNYNDQLEWE 480
VVHGKKS+++KM G +FA LR LY Y M HPGKLLFMGSEF QF ENK+ ++L+W

Sbjct: 422 VVHGKKSLLNKMFGSYEEKFAQLRLLYGYMMAHPGKLLFMGSEFAQFDEWKFABELEDWV 481

35 Query: 481 NLNDMDNQMORYTKQLNQFYKDKHKLWRIDDSFDGLEIIDADNKSETVLSFIRKDDK-G 539
+ ++++KM Y KQL YK +K + +D G E ID N +++ SFIR+ K G

Sbjct: 482 LDFELHRKMDKEYVKQLIACYKRYKPFYELDHDPRGFENIDVHNAQSIFSIFIRGKKEG 541

40 Query: 540 DLLLCVFNMTVERENFTIGVPQAGIYEEVLNTEMEFEGGVWKNHNPVTKQVATWKDYD 599
D+L+ V N T ++ + VP Y EVLN++ EFGG + +

Sbjct: 542 DVLVIVCNFTNQAYDDYKVSVPLLAPYREVLNSDAEFGGSGHVNGKRLPAFSEPFHGKP 601

Query: 600 HTLSFTLEALGASVWR 615
+ + T+P G S+ R

40 Sbjct: 602 YHVRMTIPFPFGISILR 617

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

45 Example 1423

A DNA sequence (GBSx1508) was identified in *S.agalactiae* <SEQ ID 4367> which encodes the amino acid sequence <SEQ ID 4368>. This protein is predicted to be pullulanase (pula). Analysis of this protein sequence reveals the following:

Possible site: 45
50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3194(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44685 GB:U67061 pullulanase [Bacteroides thetaiotaomicron]
Identities = 223/597 (37%), Positives = 331/597 (55%), Gaps = 55/597 (9%)
60 Query: 139 EYSETKTAFLWAPTAERVELILYHSTDETASVSKVLSMKRGTA VNYKNHKENTHGVWFT 198

-1569-

EY+ T F LW+PTA+ V L+LY + E + + M+ G G W
 Sbjct: 46 EYTPEATKFTLWSPTADEVRLMLYEA-GEGGHAYETVVKMQSGE-----EGTWTA 93
 Query: 199 ELEGNYNQAYTYRVVYRRRTFKITRDPYSIATTANGKRSIVIAPALTPEGFKISHGKE 258
 + + + YT+ V + T + A NGKR+ +I ++ P+G++ +
 Sbjct: 94 VVSKDLIGKFYTFNVKIDDKWQGDTPGINARAVGVNGKRAAIDWQSTNPDGWE----SD 149
 Query: 259 AKWRLNPNQAVIYEMHVRDFSISSETSGVKTDYHGKFKGLHQKGTVNQHGDKTTFDYVQD 318
 + L++P +IYEMH RDFS+ TSGVK GK+ L + GT+N T D++ +
 Sbjct: 150 TRPPLKSPADMIYEMHHRDFSVDSTSGVKNK--GKYLALTEHGTMNNSDKLLTGIDHLIE 207
 Query: 319 LGVNYIQLQPIFDHHTFDDDD-GHYAYNWGYDPENYNVPEASFSSNPHEPATRILELKSA 377
 LGV ++ L P FD+ + +YNWGYDP+NYNVP+ S++++P++PATR+ E K
 Sbjct: 208 LGVTHVHLPSFDYASVDETRLNENSYNWGYDEPNYNVPGSYATDPYQPATRVKEFKQM 267
 Query: 378 IQAYHDAGIGVIMDVVNHTFSSSTDSAFQLTVPDYRYRMNHNHGTQNGSGCGNETASEKE 437
 +QA H AGI VIMDVVNHTF++ +S F+ TVP Y+YR + T NGSGCGNETASE+
 Sbjct: 268 VQALHKAGIRVIMDVVNHTFNTDES NFERTVPGYFYRQKEDKTLANGSGCGNETASERL 327
 Query: 438 MCRKYILDSVLYWVKEYNIDGFRFDLMGLHDVETMNIIRNELNKIDPRILVYEGGWDMDGA 497
 M RK+++SVLYW+KEY++DGFRFDLMG+HD+ETMN IR +N +DP I +YEGGW A
 Sbjct: 328 MMRKFMVESVLYWIKYHYVDGFRFDLMGIHDIETMNEIRKAVNAVDPTTICIYEGGWAAEA 387
 Query: 498 GLTPQNK-AKKNAYQMPGIGFFNDVVRDAV---KGAEIYGEFKKGLVSGNSTEDIVAKG 553
 P + A K N Q+PG+ F+D++RD + G + G F G+ G E V G
 Sbjct: 388 PQYPADSLAMKGNIAQIPGVAVFSDDELRLDGLCGPVGDKRKGAFLAGIPGG---EMSVKFG 444
 Query: 554 ILGSDE-----LVSYI-----DPSQVLNYVEAHDNYNLNDLLWELHPNDNEKQHIYR 600
 I G+ E V+Y P Q+++YV HD L D L P+ +Q I
 Sbjct: 445 IAGAIEHPQVQCDSVNYTQKQWAKQPVQMISYVSCHDGLCLVDRLKASMPDITPEQLIRL 504
 Query: 601 VEVASAMNLLMQGMAFMQLGQEFRLRTKCYPTGDKGQLTQADKERAMNSYNAPDQVNVQVNW 660
 ++A A+ QG+ F+ G+E +R DK+ NSY +PD VN ++W
 Sbjct: 505 DKLAQAVVFTSQGIPFIYAGEEIMR-----DKQGVDSYKSPDAVNAIDW 549
 Query: 661 DNVTFHKSTINFIRKIITLKTNSPYFSYSFEEIRKHVVFESAQYHSGFISFTVEEH 717
 T + +++I L+ + P F ++RKH+ + S I+F +++H
 Sbjct: 550 RRKTTADVFMYYKRLIDLRKSHPAFRMGDAGQVRKHLEFLPVE-GSNLIAFRLKDH 605

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1424

A DNA sequence (GBSx1509) was identified in *S.agalactiae* <SEQ ID 4369> which encodes the amino acid sequence <SEQ ID 4370>. Analysis of this protein sequence reveals the following:

45 Possible site: 40
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.2368(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:CAB12492 GB:Z99107 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 151/293 (51%), Positives = 193/293 (65%), Gaps = 5/293 (1%)
 Query: 5 KKARLIYNPTSGQEIMKKNVAEVLDDILEGFGYETSAFQTTPTKNSARDEATRAAQAGFDL 64
 K+AR+IYNPTSG+EI KK++A+VL E GYETS TT A A AA FDL
 Sbjct: 2 KRARIYNPTSGREIFKKHLAQVLQKFEQAGYETSTHATT-CAGDATHAAKEAALREFDL 60
 60 Query: 65 IVAAGGDTINEVVNGIAPLKRPRKMAIPTGTNDFARALKIPRGNPIEATKLIGKNQI 124
 I+AAGGDTINEVVNG+APL RP + +IP GTTND FARAL IPR + ++A +

-1570-

Sbjct: 61 IIAAGGDTINEVVNGLAPLDNRPTLGVIPVGTINDFARALGIPREDILKAADTVINGVA 120

Query: 125 VKMDIGQAQEDNYFINIAAAGSLTELTYSVPSQLKTTFGYLAYLAKGVELLPRVRKVPVK 184
+DIGQ YFINIA G LTELTY VPS+LKT G LAY KG+E+LP +R V+

5 Sbjct: 121 RPIDIGQVN-GQYFINIAGGRLTELTYDVPSKLKTMGLQLAYYLKGMEMLPRLPTEVE 179

Query: 185 ITHDKGEFIGDASMIFVAITNSVGGFEQIAPDAKLDGKFTLILVKTANLIEIMHLIRLV 244
I +D F G+ + V +TNSVGGFE++APD+ L+DG F L+++K ANL E + + +

10 Sbjct: 180 IEYDGKLFQGEIMLFLVLTNSVGGFEKLAPDSSLDGMFDLMILKKANLAEFIRVATMA 239

Query: 245 LAGGKHINDKRVEYIKTSYLTIEPLSDERMMINLDGEYGGDAPITLANLKNHI 297
L G+HIND+ + Y K + + + E+M +MLDGEYGG P NL HI

Sbjct: 240 LR-GEHINDQHIIYTKANRVKVN--VSEKMLNLDGEYGGMLPGEFVNLYRHI 289

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4371> which encodes the amino acid sequence <SEQ ID 4372>. Analysis of this protein sequence reveals the following:

Possible site: 40
>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2501(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 272/334 (81%), Positives = 300/334 (89%)

Query: 1 MKKQKKARLIYNPTSGQEIMKKNVAEVLIDILEGFGYETSAFQTPTTKNSARDEATRAAQA 60
MKKQ +ARLIYNPTSGQE+M+K+V EVLDILEGFGYETSAFQTT KNSA +EA RAA+A

30 Sbjct: 1 MKKQLRARLIYNPTSGQELMRKSVPEVLIDILEGFGYETSAFQTTAQKNSALNEARRAAKA 60

Query: 61 GFDLIVAAGGDTINEVVNGIAPLKRPKMAIIPTGTTNDFARALKIPRGNPPEATKLIG 120
GFDL++AAGGDTINEVVNGIAPLK+RPKMAIIPTGTTNDFARALK+PRGNP +A KLIG

35 Sbjct: 61 GFDLLIAAGGDTINEVVNGIAPLKKRPKMAIIPTGTTNDFARALKVPRGNPSQAACLIG 120

Query: 121 KNQIVKMDIGQAQEDNYFINIAAAGSLTELTYSVPSQLKTTFGYLAYLAKGVELLPRVRK 180
KNQ ++MDIG+A++D YFINIAAAGSLTELTYSVPSQLKT FGYLAYLAKGVELLPRV

Sbjct: 121 KNQTIQMDIGRAKDKTYFINIAAAGSLTELTYSVPSQLKTMFGYLAYLAKGVELLPRVSN 180

40 Query: 181 VPKVITHDKGEFIGDASMIFVAITNSVGGFEQIAPDAKLDGKFTLILVKTANLIEIMHL 240
VPKVITHDKG F G SMIF AITNSVGGFE IAPDAKLDG FTLIL+KTANL EI+HL

Sbjct: 181 VPKVITHDKGVFEGQVSMIFAATNSVGGFEMIAPDAKLDGDMFTLILIKTANLFEIVHL 240

45 Query: 241 IRLVLGCKHINDKRVEYIKTSYLTIEPLSDERMMINLDGEYGGDAPITLANLKNHIRFF 300
+RL+L GKKHI D+RVEYIKTS + IEP +RMMINLDGEYGGDAPITL NLKNHI FF

Sbjct: 241 LRLILDGCKHITDRRVEYIKTSKIVIEPQCGKRMMINLDGEYGGDAPITLENLKNHITFF 300

Query: 301 ANTDEISDDALVLDKDELAIEATAQKFANEVDDL 334
A+TD ISDDALVLD+DEL IE I +KFA+EV+DL

50 Sbjct: 301 ADTDLISDDALVLDQDELEIEEIVKKFAHEVEDL 334

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1425

55 A DNA sequence (GBSx1510) was identified in *S.agalactiae* <SEQ ID 4373> which encodes the amino acid sequence <SEQ ID 4374>. This protein is predicted to be DNA ligase (ligA-1). Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

-1571-

INTEGRAL Likelihood = -0.27 Transmembrane 363 - 379 (363 - 379)

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9763> which encodes amino acid sequence <SEQ ID 9764> was also identified.

- 10 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12482 GB:Z99107 similar to DNA ligase [Bacillus subtilis]
 Identities = 346/657 (52%), Positives = 462/657 (69%), Gaps = 8/657 (1%)

- 15 Query: 2 ENRMNELVSLNQYAKYYTQDNPTVSDSQYDQLYRELVELEKQHPENILPNSPTHRVGG 61
 + R EL +N+Y+ EYYT D P+V D++YD+L +EL+ +E++HP+ P+SPT RVGG
 Sbjct: 7 KQRAEELRRTINKYSYEEYTLDEPSVPAEYDRMQELIAIEEHPDLRTPDSPTQRVGG 66
- 20 Query: 62 LVLEGFEKYQHEYPYLSLQDAFSKEELIAFDKRVKAEF-PTAAYMAELKIDGLSVSLTYV 120
 VLE F+K H P+ SL +AF+ ++L FD+RV+ AY ELKIDGL+VSL Y
 Sbjct: 67 AVLEAFQKVTHGTPMLSLGNAFNADDLDRDFDRRVQSVGDDVAYNVELKIDGLAVSLRYE 126
- 25 Query: 121 NGVLQVGATRGDGNIGENITENLKRVDHDIPLHLDQSLDITVRGECYLPKESFEAINIEKR 180
 +G GATRGDG GE+ITENLK + +IPL +++ L I VRGE Y+PK SFEA+N E+
 Sbjct: 127 DGYFVRGATRGDGTGEDITENLKTIRNIPLKMNRELSIEVRGEAYMPKRSFEALNEERI 186
- 30 Query: 181 ANGEQEFANPRNAAAGTLRQLNTGIVAKRKLATFLYQEASPTQK--ETQDVLKELESYG 238
 N E+ FANPRNAAAG+LRQL+ I AKR L F+Y A + ETQ L L+ G
 Sbjct: 187 KNEEBPFANPRNAAAGSLRQLDPKIAAKRNLDIFVYSIAELDEMGVETQSQGLDFLDELG 246
- 35 Query: 299 IAYKFPAAEKEAEILSVDWTVGRTGVVTPANTLPVQLAGTTVSRATLHNVYIAEKDIR 358
 IAYKFPAAE ++L ++ VGRGTG+TPTA L PV++AGTTVSRA+LEN D I EKDIR
 Sbjct: 307 IAYKFPAAEEVVTKLLDIELNVGRTGVITPTAILEPVKVAGTTVSRASLHNEELIKEKDIR 366
- 40 Query: 359 IGDTVVVYKAGDIIPAVLNVVMSKRNQQEVM-L-IPKLCPSGSELVHFEGEVALRCINPL 417
 I D VVV KAGDIIP V+NV++ +R +E +P CP CGSELV EGEVALRCINP
 Sbjct: 367 ILDKVVVKAGDIIPVVNVLDVQRTGEEKEFSMPTECECGSELVRIEGEVALRCINPE 426
- 45 Query: 418 CPNQIKERLAHFASRDAMNITGFGPSLVEKLFDAHLIADVADIYRLSIENLLTLDGIKEK 477
 CP QI+E L HF SR+AMNI G G ++ +LF+ +L+ +VAD+Y+L+ E ++ L+ + EK
 Sbjct: 427 CPAQIREGLIHVSRNAMNIDGLGERVITQLFEENLVRNVADLYKLTKEVQLERMGEK 486
- 50 Query: 478 SATKIYHAIQSSKENSAAEKLFLGIRHVGSKASRLLEEFGNLRQLSQASQESIASIDG 537
 S + +IQ SKENS E+LLFGLGIR +GSKA++ L F +L L +AS+E + ++D
 Sbjct: 487 STENLISSIQSKENSLERLLFGLGIRFIGSKAAKTLAMHFESLENLKKASKEELLAVDE 546
- 55 Query: 538 LGGVIAKSLHTFFEKEEVDKLLLEELTSYNNVFNVLG---KRVSTDAQLSGLTVVLTGKL 593
 +G +A ++ T+F KEE+ +LL EL VN Y G K +D+ +G T+VLTGKL
 Sbjct: 547 IGEKMADAVITYFHKEMLELLNELQELGVNTLYKGPVKVKAEDSDSYFAGKTTVLTGKL 606
- Query: 594 EKMTRNEAKEKLQNLGAKVTGSVSKKTDLIVAGSDAGSKLTKAQDLGITIQDEDWLL 650
 E+++RNEAK +++ LG K+TGSVSK TDL++AG AGSKLTKAQ+L I + +E+ L+
 Sbjct: 607 EELSRNEAKAQIEALGGKLTGSVSKNTDLVIAGEAAGSKLTKAQELNIEVWNEEQLM 663

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4375> which encodes the amino acid sequence <SEQ ID 4376>. Analysis of this protein sequence reveals the following:

- 60 Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.43 Transmembrane 363 - 379 (363 - 379)

-1572-

----- Final Results -----

bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 472/652 (72%), Positives = 556/652 (84%)

10 Query: 1 MENRMNELVSLINQYAKYYTQDNPTVSDSQYDQLYRELVELEKQHPENILPNSPTHRVG 60
 M+ R+ EL LLN+Y +YYT+D P+VSDS YD+LYRELV LE+ +PE +L +SPT +VG
 Sbjct: 1 MKKRIKELTDLLNRYRYDYTTKDAPSVSDSDYDKLYRELVTLEQSYPEYVLQDSPTQQVG 60

15 Query: 61 GLVLEGFEEKYQHEYPYLSLQDAFSKEELIAFDKRVKAEFPTAAYMAELKIDGLSVSLTYV 120
 G +L+GFEEKY+H+YPL+SLQDAFS+EEL AFDKRVKAEFP A Y+AELKIDGLS+SL+Y
 Sbjct: 61 GTILKGFEKYRHOYPLFSLQDAFSREELDAFDKRVKAEFPNATYLAELKIDGLSISLSYE 120

20 Query: 121 NGVLQVGATRGDGNIGENITENLKRVDIPLHLDQSLDITVRGECYLPKESFEAINIEKR 180
 NG LQVGATRGDGNIGENITEN+K++ DIP L + L ITVRGE Y+ ++SF+AIN ++
 Sbjct: 121 NGFLQVGATRGDGNIGENITENIKKIDIPYQLSEPLTTITVRGEAYMSRQSFKAINEARQ 180

25 Query: 181 ANGEQEFANPRNAAAGTLRQLNTGIVAKRKLATFLYQEASPTQKETQDDVLKELESYGFS 240
 NGE EFANPRNAAAGTLRQL+T +VAKR+LATFLYQEASPT + Q++VL EL GFS
 Sbjct: 181 ENGETEFANPRNAAAGTLRQLDTSVVAKRQLATFLYQEASPTARNQONEVLAEADLGF 240

30 Query: 241 VNHHRLISSMEKIWDFITQIEKDRVSLPYDIDGIVIKVNSIAMQEELGFTVKAPRWAI 300
 VN + ++SSM++IWDFT+TIE R L YDIDG+VIKVNS+AMQEELGFTVKAPRWAI
 Sbjct: 241 VNPYYQLTSSMDEIWDFIKTIEAKRDQLAYDIDGVVIKVNSLAMQEELGFTVKAPRWAI 300

35 Query: 301 YKFPAAEKEAEILSVDWTVGRTGVVTPPTANLTPVQLAGTTVSRATLHNVDYIAEKDIRIG 360
 YKFPAAEKEAEILSVDWTVGRTGVVTPPTANLTPVQLAGTTVSRATLHNVDYIAEKDIRIG
 Sbjct: 301 YKFPAAEKEAEILSVDWTVGRTGVVTPPTANLTPVQLAGTTVSRATLHNVDYIAEKDIRIG 360

40 Query: 361 DTVVVYKAGDIIPAVLNVVMSKRNQEVMLIPKLCPCGSELVHFEGEVALRCINPLCPN 420
 DTV+VYKAGDIIPAVLNVVMSKRNQEVMLIPKLCPCGSELVHFE EVALRCINPLCP+
 Sbjct: 361 DTVIVYKAGDIIPAVLNVVMSKRNQEVMLIPKLCPCGSELVHFEDEVALRCINPLCPS 420

45 Query: 421 QIKERLAHFASRDAMNITGFGPSLVEKLFDAHLIADVADIYRLSIENLLTLDGIKEKSAT 480
 I+ L HFASRDAMNITG GP++VEKLF A + DVADIY+L+ E+ + LDGIKEKSA
 Sbjct: 421 LIQRSLEHFASRDAMNITGLGPAIVEKLFAGFVHDVADIYQLTKEDFMQLDGIKEKSAD 480

50 Query: 481 KIYHAIQSSKENSAAEKLLFGLGIRHVGSKASRLLEEFGNLRQLSQASQESIASIDGLG 540
 K+ AI++SK NSAEKLLFGLGIRH+GSK SRL+LE +G++ L A +E IA IDGLG
 Sbjct: 481 KLLAAIEASKSNSAEKLLFGLGIRHIGSKVSRILEVYGDISALLTAKEEETIARIDGLGS 540

55 Query: 541 VIAKSLHTFFEKEEVDKLEELTSYNVNFNYLGKRVSTDAQLSGLTVVLTGKLEKMTNE 600
 IA+SL +FE++ L++EL + VN +Y G++V++DA L GLTVVLTGKL ++ RNE
 Sbjct: 541 TIAQSLTQYFEQKTAAILVDELKTAGVMNHYSQGKVNDAALFGLTVVLTGKLNQLNRNE 600

Query: 601 AKEKQLNLGAKVTGSVSKKTDLIVAGSDAGSKLTKAQDLGITIQEDDWLLNL 652
 AK+KL+ LGAKVTGSVSKKTDL++AGSDAGSKL KA+ LGI I+DEDWL L
 Sbjct: 601 AKDKLEALGAKVTGSVSKKTDLVIAGSDAGSKLEKAKSLGIRIEDEDWLRQL 652

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1426

A DNA sequence (GBSx1511) was identified in *S.agalactiae* <SEQ ID 4377> which encodes the amino acid sequence <SEQ ID 4378>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have a cleavable N-term signal seq.

60 INTEGRAL Likelihood = -5.63 Transmembrane 110 - 126 (108 - 128)
 INTEGRAL Likelihood = -2.13 Transmembrane 142 - 158 (141 - 159)
 INTEGRAL Likelihood = -1.12 Transmembrane 75 - 91 (75 - 93)

-1573-

----- Final Results -----

5 bacterial membrane --- Certainty=0.3251(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA68244 GB:X99978 citrulline cluster-linked gene [Lactobacillus
 plantarum]
 10 Identities = 56/158 (35%), Positives = 91/158 (57%), Gaps = 8/158 (5%)
 Query: 13 AIVTAIYIVLITITPPFNIAIYGAQFRVSEMLNFLAFYHRKYLFAVTLGCMISNLYSFG- 71
 A+V A+Y+VL + P ++A GA QFRVSE LN LA ++RKY++ + G ++ + + G
 15 Sbjct: 13 ALVAAMYVVLCLGPAAFSLASGAIQFRVSEGLNHLAVFNRYKIWGIVAGVILFDAGPGA 72
 Query: 72 -MIDVFVGGGSTALLFVYLGTILFKQYQKDYLFNGLINKAFFFFSFFFAASMITVAVELKI 130
 +++V GGG +LL + + T L + K L+N A F S F A MIT+ +
 Sbjct: 73 SLLNVLFGGQSLLALLVLTWLAPKL-KTVWQRMMLNIALFTVSMFMIALMITM-----M 126
 20 Query: 131 VAGLPILLTWTTLTAVGELASLLVGAVLVDKLSRHVDFT 168
 +G+ T+LTTA+ EL + + A ++ L R + F+
 Sbjct: 127 SSGVAFWPTYLTALTSELIIMSITAPIMYSLDVRLHFS 164

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4379> which encodes the amino acid
 sequence <SEQ ID 4380>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 30 INTEGRAL Likelihood = -4.41 Transmembrane 75 - 91 (70 - 94)
 INTEGRAL Likelihood = -3.82 Transmembrane 12 - 28 (8 - 28)
 INTEGRAL Likelihood = -2.28 Transmembrane 141 - 157 (140 - 158)
 INTEGRAL Likelihood = -0.64 Transmembrane 110 - 126 (110 - 126)
 INTEGRAL Likelihood = -0.59 Transmembrane 55 - 71 (54 - 73)

----- Final Results -----

35 bacterial membrane --- Certainty=0.2763(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 114/167 (68%), Positives = 137/167 (81%), Gaps = 1/167 (0%)
 Query: 1 MNTFTTRDYAHMAIVTAIYIVLITITPPFNIAIYGAQFRVSEMLNFLAFYHRKYLFAVTL 60
 M T DY H+ +V A+Y+VLITITPP NAI+YG YQFR+SEM+NFLAFYHRKY+ AVTL
 45 Sbjct: 1 MTKLTVHDYVHIGLVAALYVVLITITPPLNAISYGYQFRISEMMNFLAFYHRKYIIAVTL 60
 Query: 61 GCMISNLYSFGMIDVFVGGGSTALLFVYLGTILFKQYQKDYLFNGLINKAFFFFSFFFAAS 120
 GCMI+N YSFG+IDVFVGGGSTL+FV LG ILP +YQKDYLFNG+ NKAF +FSFFFA S
 Sbjct: 61 GCMIANFYSPGLIDVFVGGGSTALLFVTLGVILFSKYQKDYLFNGIFNKAFVYFSFFFATS 120
 50 Query: 121 MITVAVELKIVAGLPILLTWTTLTAVGELASLLVGAVLVDKLSRHVD 167
 M VA+EL G P LLTW TTA+GEL SLL+G+++DKLS+ + F
 Sbjct: 121 MFNVAIELYFF-GAPFLLTWFTTALGELVSLIGSLIIDKLSQRISF 166

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1427

A DNA sequence (GBSx1513) was identified in *S.agalactiae* <SEQ ID 4381> which encodes the amino
 acid sequence <SEQ ID 4382>. Analysis of this protein sequence reveals the following:

-1574-

Possible site: 53

>>> Seems to have no N-terminal signal sequence

5 INTEGRAL Likelihood = -11.20 Transmembrane 255 - 271 (245 - 281)
 INTEGRAL Likelihood = -10.72 Transmembrane 141 - 157 (132 - 165)
 INTEGRAL Likelihood = -8.17 Transmembrane 189 - 205 (185 - 208)
 INTEGRAL Likelihood = -7.01 Transmembrane 36 - 52 (33 - 60)

----- Final Results -----

10 bacterial membrane --- Certainty=0.5479(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC35915 GB:AF071085 Orfde2 [Enterococcus faecalis]
 Identities = 83/276 (30%), Positives = 157/276 (56%), Gaps = 3/276 (1%)

 Query: 17 RPIQVFMRFQSAEMDLSAIAVAYYLLVTAFFLLVIAANIFPYFHINVSDDLMLQKNLP 76
 R I+ H +AE+ S++ VAYYLL++ FPLL+ N+ PY I+ + +L + + +P
 Sbjct: 15 RFIETTQSHMVTAEIGNSSVVVAYYLLLSLFPLLIAVGNVLPYLRIDPNSVLPYIAEAIP 74

 Query: 77 KNIYEPASRLAVDAFSKPSTGILGFASLTAFWTMSKSLTSLQKAINKAYGVDQHRDFVIS 136
 K++Y+ ++ S G+L ++L AFW+ S+S+ +LQ A+NKA+GV+Q ++F++
 Sbjct: 75 KDVYKNLEPAIRSLLTQRSGGLLSVSALAAFWSASQSINALQNAMNKAFGVEQRKNFILV 134

 Query: 137 RLVGVGTGLIILFLTLFVLIFSTFSPVQLIIVNMYDLGDTLTAWLLNLAQPVTFLTIFL 196
 R+V L+ + + V++ + +++++ ++ L P+T + + +
 Sbjct: 135 RVVSFLVILFEMVAIVGVVILGLGQYIIELLQPIFHYSTSVIDTFQALKWPLTTVVLLV 194

 Query: 197 GIGILYFILPNARIRKVRYVPGTLFSTFVIGFFSNLISQYVLNRVEKMVDIKTFGSVVI 256
 + ++Y ++PN ++ +R ++PG +FST S + YV ++ + GS +
 Sbjct: 195 IMCLIIYAVVNRKL-SLRSILPGAI FSTVGWMLLSQIFGLYVKYFSSRIASVQIIGSFI- 252

 Query: 257 FILMLWFIFLAHIMILGAILNASVQEIATGKIESRR 292
 ILMLW F A I+ILGAI+NA V E G E ++
 Sbjct: 253 -ILMLWLNFAATIIILGAIVNAVVDLYLXGKKEKKQ 287

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4383> which encodes the amino acid sequence <SEQ ID 4384>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

40 INTEGRAL Likelihood = -12.58 Transmembrane 141 - 157 (132 - 168)
 INTEGRAL Likelihood = -12.15 Transmembrane 189 - 205 (177 - 210)
 INTEGRAL Likelihood = -11.68 Transmembrane 256 - 272 (245 - 280)
 INTEGRAL Likelihood = -7.54 Transmembrane 36 - 52 (33 - 60)

----- Final Results -----

50 bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 >GP:CAA68244 GB:X99978 citrulline cluster-linked gene [Lactobacillus
 plantarum]
 Identities = 53/170 (31%), Positives = 92/170 (53%), Gaps = 11/170 (6%)

 Query: 1 MTKLTVHDYVHIGLVAALYVVLTTITPPLNAISYGMVQFRISEMMNFLAFYHRKYIIAVTL 60
 MT+ + ++ LVAA+YVVL + P +++ G QFR+SE +N LA ++RKYI +
 Sbjct: 1 MTQSKIRPWIINALVAAMYVVLCLGPAAFSLASGAIQFRVSEGLNHLAVFNRYKIWGIVA 60

 Query: 61 GCMIANFYSG--LIDVFVGGGSLTIFVTILGVILFSKYQKDYLFNGIFNKAFVYFSFFFA 118
 G ++ + + G L++V GGG +L+ + + L K + ++ + + + F
 Sbjct: 61 GVILFDAFGPGASLLNVLFGGGQSLALLLVLTWLA PKLKT-----VWQRMLLNIA-LFT 113

 Query: 119 TSMFNVA--IELYFFGAPFLLTWFTTALGELVSLIGSLIIDKLSQRISF 166

-1575-

SMF +A I + G F T+ TTAL EL+ + I + I+ L + + F
 Sbjct: 114 VSMFMIALMITMSSGVAFWPTYLTALSELIIIMSITAPIMYSLDRVLHF 163
 !GB:AF071085 Orfde2 [Enterococcus faecalis] 176 2e-43

5 >GP:AAC35915 GB:AF071085 Orfde2 [Enterococcus faecalis]
 Identities = 90/271 (33%), Positives = 155/271 (56%), Gaps = 3/271 (1%)

Query: 19 IQVFMRLHLSAEMDLSAIAVAYYLILTAFLPLIVIAANIFPYLNIDILLRLMKQNLPKD 78
 I+ H+ +AE+ S++ VAYYL+L+ FPL++ N+ PYL ID +L + + +PKD
 10 Sbjct: 17 IETTQSHMVTAEIGNSSVVVAYYLLLSLFPLLIAGVNLVPLRIDPNSVLPYIAEAIPKD 76

Query: 79 IFRPASAIVENIFSKPSGVLGVATLTGLWTMSRSLTSLQKAINKAYGASQHRDFFIGHL 138
 +++ + ++ ++ SG +L V+ L W+ S+S+ +LQ A+NKA+G Q ++F + +
 15 Sbjct: 77 VYKNLEPAIRSLLTQRSGLLSVSALAFAFWSAQSQINALQNAMNKAFGVBQRKNFILVRV 136

Query: 139 VGLLTSLLILFLLAFALIFSIFSKAAIQVLDKHYHLSDNITTIFLLLIQIPITVLIIFVGL 198
 V L L+ + + ++ + I++L +H S ++ F L P+T +++ V +
 20 Sbjct: 137 VSFLVILLFMVAIVGVVILGLGQYIIELLQPIFYHSTSVIDTFQALKWPLTTVVLLVIM 196

Query: 199 MLLYFLLPNVKIKIRYILPGTLFTSFVMTFLSNLVGNVYVYVVERMVDIKMFGSVMIFI 258
 L+Y ++PN K+ +R ILPG +F++ LS + G YV Y R+ ++ GS I
 25 Sbjct: 197 CLIYAVVPNRKL-SLRISILPGAIFSTVGWMLLSQIFGLYVKYFSSRIASYQIIGS--FII 253

Query: 259 IMLWFIFLARILILGAIFNATYQEMSLGKLE 289
 +MLW F A I+ILGAI NA E G E
 25 Sbjct: 254 LMLWLNFAATIIILGAIVNAVVDLYLXGXKE 284

An alignment of the GAS and GBS proteins is shown below.

Identities = 188/302 (62%), Positives = 244/302 (80%)

30 Query: 1 MKLKKFFEDLLAKLEYRPIQVFMRFQSAEMDLSAIAVAYYLLVTAFLPLIVIAANIFPYF 60
 M KK+F+ +L+K +Y PIQVFMRLHLSAEMDLSAIAVAYYL++TAFPL+VIAANIFPY
 Sbjct: 1 MAEKKWFDPKVLKQYEPPIQVFMRLHLSAEMDLSAIAVAYYLILTAFLPLIVIAANIFPYL 60

35 Query: 61 HINVSDDLMLQKQNLKPNYEPASRLAVDAFSPSTGILGFASLTAFWTMSKSLTSLQKA 120
 +I+++DLL LM++NLPK+I+ PAS + + FSKPS +LG A+LT WTMS+SLTSLQKA
 Sbjct: 61 NIDIADLLRLMKQNLPKDIFRPASAIVENIFSKPSGVLGVATLTGLWTMSRSLTSLQKA 120

40 Query: 121 INKAYGVDQHRDFVISRLVGVGTGLIILFLLTFVLIFSTFSKPVLQIIVNMYDLGDTLTA 180
 INKAYG QHRDF I LVG+ T LIILFLL F LIFS FSK +Q++ Y L D +T
 Sbjct: 121 INKAYGASQHRDFFIGHLVGLLTSLLILFLLAFALIFSIFSKAAIQVLDKHYHLSDNITT 180

45 Query: 181 WLLNLAQPVTFLITFLGIGILYFILPNARIRKVRYPVPGTLFSTFVIGFFSNLISQYVIN 240
 L L QP+T L IF+G+ +LYF+LPN +I+K+RY++PGTLF++FV+ F SNL+ YV+
 Sbjct: 181 IFLLLIPIITVLIIFVGLMMLLYFLLPNVKIKIRYILPGTLFTSFVMTFLSNLVGNVYVY 240

50 Query: 241 RVEKMVDIKTFGSVVIFILMLWFIFLAHIMILGAILNASVQEIATGKIESRRGDIMSLIQ 300
 VE+MVDIK FGSV+IFI+MLWFIFLA I+ILGAI NA+ QE++ GK+E R GD+++++
 Sbjct: 241 NVERMVDIKMFGSVMIFIIMLWFIFLARILILGAIFNATYQEMSLGKLEGRSGDMIAILK 300

55 Query: 301 KS 302
 K+
 Sbjct: 301 KT 302

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1428

A DNA sequence (GBSx1514) was identified in *S.agalactiae* <SEQ ID 4385> which encodes the amino acid sequence <SEQ ID 4386>. Analysis of this protein sequence reveals the following:

60 Possible site: 22
 >>> Seems to have no N-terminal signal sequence

-1576-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4200(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 1429

A DNA sequence (GBSx1515) was identified in *S.galactiae* <SEQ ID 4387> which encodes the amino acid sequence <SEQ ID 4388>. This protein is predicted to be methionine aminopeptidase (map). Analysis of this protein sequence reveals the following:

Possible site: 14

15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2342(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9761> which encodes amino acid sequence <SEQ ID 9762> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAC35914 GB:AF071085 methionine aminopeptidase A [Enterococcus
 faecalis]
 Identities = 101/207 (48%), Positives = 128/207 (61%), Gaps = 31/207 (14%)

30 Query: 1 MITLKSAREIEAMDRAGDFLASIHIGLRDIKPGVDMWEVEEYVRRRCKEENVLPLQIGV 60
 MITLKS REIE MD +G+ LA +H LR IKPG+ W++E +VR + + QIG
 Sbjct: 1 MITLKS PREIEMMDES GELLADVHRHLRTFIKPGITSWDIEVFVRDFIESHGGVAAQIGY 60

35 Query: 61 DGA VM DY P YATCCGLNDEVAHAFPRHYTLKQGDLLKVDMLSEPLDKSIVDVSSLNFDNV 120
 +G Y YATCC +NDE+ H FPR LK GDL+KVDM +
 Sbjct: 61 EG----YKYATCCSINDEICHGFPKKVLDKGDLLKVDMLSEPLDKSIVDVSSLNFDNV 98

40 Query: 121 AQMKKYTETYSGLADSCWAYAVGEVSVQEVKDLMSVTREAMYIGIEKAVIGNRIGDIGAA 180
 G ++DSCW+Y VGE + E+ LM VT++A+Y+GIE+A +GNRIGDIG A
 Sbjct: 99 -----LKG AISDSCWSYVVGESTPEIDRLMEVTKKALYLGIEQAQVGNRIGDIGHA 149

45 Query: 181 IQDYAESRGYGVVRDLVGHGVGPTMHE 207
 IQ Y E GYGVVRD VGHG+GPT+HE
 Sbjct: 150 IQTYVEGEGYGVVRDFVGHGIGPTIHE 176

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4389> which encodes the amino acid sequence <SEQ ID 4390>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2082(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below.

-1577-

Identities = 256/286 (89%), Positives = 273/286 (94%)

Query: 1 MITLKSAREIEAMDRAGDFLASIHIGLRDIIKPGVDMWEVEEYVRRRCKEENVLPQIGV 60
 5 Sbjct: 1 MITLKSAREIEAMDRAGDFLAGIHIGLRDIIKPGVDMWEVEAYVRRRCKEDNVLPQIGV 60

Query: 61 DGAVMDYPYATCCGLNDEVAHAFPRHYTLKQGDLLKVDMLSEPLDKSIVDVSSLNFDNV 120
 10 Sbjct: 61 DGHMMDYPYATCCGLNDEVAHAFPRHYILKEGDLLKVDMLSEPLDKSIVDVAALDFDNV 120

Query: 121 AQMKKYTETYSGLADSCWAYAVGEVSQEVKDLMSVTREAMYIGIEKAVIGNRIGDIGAA 180
 Sbjct: 121 PEMKKWTGYSYTGGLADSCWAYAVGTPSDEIKQLMDVTKEAMYRGIEKAVIGNRIGDIGAA 180

15 Query: 181 IQDYAESRGYGVVRDLVGHGVGPTMHEEPMVPNYGTAGRGLRLREGMVLTIETPINTGTW 240
 Sbjct: 181 VQEYAESFGYGVVRDLVGHGVGPTMHEEPMVPNYGTAGRGLRLKEGMVLTIETPINTGTW 240

20 Query: 241 EIDTDMKTGWAHKTLDGGLSCQYEHQFVITKDGVPILTSCQGEERTY 286
 Sbjct: 241 EIDTDKTGWAHKTLDGGLSCQYEHQFVITKDGVPILTSCQGEERTY 286

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1430

A DNA sequence (GBSx1516) was identified in *S.agalactiae* <SEQ ID 4391> which encodes the amino acid sequence <SEQ ID 4392>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3473(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9759> which encodes amino acid sequence <SEQ ID 9760> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06894 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 40 Identities = 158/431 (36%), Positives = 270/431 (61%), Gaps = 6/431 (1%)

Query: 6 SKHQEILEYLENLAVGKRVSVRSISNHLKVS DGTAYRAIKEAENRGIVETRPRSGTVRVA 65
 +KH++IL+Y+ NL VG+++SVR I+ L+VS+GTAYRAIKEAEN+G+V T R GT+R+
 45 Sbjct: 3 TKHEQILQYITNLEVGEKISVRRIAKDLQVSEG TAYRAIKEAENQGLVSTIERVGTIRIE 62

Query: 66 QKAKVNIEKLTAYAEIARISDSQVVGIEGLSKEFSKFSIGAMTHRNIEKYL VQGGLLIVG 125
 +K K NIEKLTAYE+ I D QV+ G +GL K ++F IGAM + +Y+ G LLIVG
 Sbjct: 63 KKQKENIEKLTAYAEVVNI VDGQVLGGRDGLHKT LNR FVIGAMKLDAMRYVEPGNLLIVG 122

50 Query: 126 DRDEIQHLALQHQNAILVTGGFNVSPVCRLADKLQIPVMVTHYDTFTVSTMINHTLSNA 185
 +R ++ +AL+ A+L+TGGF+ S +LAD+L +PV+ T YDTFTV+TMIN + +
 Sbjct: 123 NRYQVHQIALEAGAAVLITGGFDTSDEAIKLADELDPVISTSYDTFTVATMINRAIDQ 182

55 Query: 186 KIRTDLKTVEQVYQSQMDYGF LAQDDTVKEFNLLVKQTKNVRFPVQNANVVGVVSVQD 245
 I+ ++ V+ + D ++ ++ V +++ L ++T + R+P++++ + G+V+ +D
 Sbjct: 183 LIKKEITLVDDILIPLDQTYMTTENVVGKWHELNEKTGHSRYPVIDENMKIQGMVAAKD 242

Query: 246 ILGKDKEVKLATVMSKNIIVAKPRMSLANISQKMIFEDLNMPVVSDDFELLGVITRRQA 305
 +L + + VM+KN I R S+A ++ M++E + ++PV+ +L+GV++R+
 60 Sbjct: 243 VLNASRHTPIEKVMTKNPITVSERTSVA AVAHVMVWEGIELLPVIDSHRKLIGVVSQDQV 302

-1578-

5 Query: 306 VENLSMSQ-----GTDLYTYSQILSNLQIEDG-HFSFLVEPAMIDHTGSLTQGVLTFL 359
 ++ L M Q G + L+ + G + + P M + G+++ GV+T +
 Sbjct: 303 LKALQMIQRQPHVGETIEDLMTINGLNESSDQGDSEYEVITPQMTNQLGTISHGVMTSLV 362

Query: 360 KEICIRVLTRKHQRSIVVKQMTLYFLQPVQIDEIIMVTPTIIEKRREATLDLELKLENK 419
 E RVL + + +VV+ +TLYFL+PVQID + + P ++ R+ +D+E+ E +
 Sbjct: 363 IESGSRVLRKYKKGDLVVENITLYFLKPVQIDSRLTIRPRVLEIGRKHGKIDVEMYHEGE 422

10 Query: 420 IIAKAMIAVKI 430
 I+ KA+ +I
 Sbjct: 423 IVGKALFMAQI 433

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4393> which encodes the amino acid
 sequence <SEQ ID 4394>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3011(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 267/431 (61%), Positives = 351/431 (80%)

Query: 1 MIIIVMSKHQEIILEYLENLAVGKRVSRSISNHLKVS DGTAYRAIKEAENRGIVETRPRSG 60
 +II+MSKHQ+IL+YLE LA+GK+VSVRSISNHLKVS DGTAYRAIKEAENRGIVET+PRSG
 Sbjct: 1 VIIIVMSKHQDILDYLEKLAIGKVSRSISNHLKVS DGTAYRAIKEAENRGIVETKPRSG 60

30 Query: 61 TVRVAQAKAKVNIEKLTYAEIARISDSQVVGIEGLSKEFSKFSIGAMTHRNIEKYLVOGG 120
 TVR+ +K +V I++LTY+EIARISDS+V+AG GL EFS+FSIGAMT +NI +YLV+GG
 Sbjct: 61 TVRIEKKGRVRIDRLTYSEIARISDSSEVLGAGLGHFSRFSIGAMTQQNIRRYLVKGG 120

35 Query: 121 LLIVGDRDEIQHLALQHQNAILVTGGFNVSPVCR LADKLQIPVMVTHYDTFTVSTMINH 180
 LLIVGDR+ IQ LAL++ NAILVTGGF VS V +A+ +IPVMVTHYDTFTV+TMINH
 Sbjct: 121 LLIVGDRETIQLLALENHNAILVTGGFPVSKRVIEMANNQRI PVMVTHYDTFTVATMINH 180

40 Query: 181 TLSNAKIRTDLKTVEQVYQSQMDYGF LAQDDTVKEFNLLVKQTKNVRFPIVNOANVVVG 240
 LSN +I+TDLKTVEQV DYG+L +D +V+EFN L+K+T+ VRFP+++ V+GV
 Sbjct: 181 ALSNIRIKTDLKTVEQVMIPITDYG YLCESSVEEFNTLIKKTRQVRFPVLDYKRVIGV 240

45 Query: 241 VSVQDILGKDKEVKLATVMSKNIIIVAKPRMSLANISQKMIFEDLNMPVVSDDFELLGVI 300
 VS++D++ + KL VMSKN I A+P SLANISQKMIFEDLN M+PV ++ LLG+I
 Sbjct: 241 VSMRDVVDQLPTTKLTKVMSKNPITARPNTSLANISQKMIFEDLNMLPVTDEENNLGMI 300

50 Query: 301 TRRQAVENLSMSQGTDLTYTYSQILSNLQIEDGHFSFLVEPAMIDHTGSLTQGVLTFL 360
 TRRQA+ENL Q + YTYS+QILSNL+ ++ +VEP MID G+++ GV++EFLK
 Sbjct: 301 TRRQAMENLNHQPNNPYTYSEQILSNLEETVDYYQVVVEPTMIDSAGNMSGVISEFLK 360

Query: 361 EICIRVLTRKHQRSIVVKQMTLYFLQPVQIDEIIMVTPTIIEKRREATLDLELKLENKI 420
 EI IR LT+KHQ++I+++QM +YFL +QI++ + + P II+E RR +T+D+E+ +++++
 Sbjct: 361 EISIRALTKKHQKNIIIEQMMVYFLHAIQIEDELKIYPKIITENRRSSTIDIEIFVDDQV 420

55 Query: 421 IAKAMIAVKIN 431
 IAKA+I KIN
 Sbjct: 421 IAKAIITTKIN 431

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

-1579-

Example 1431

A DNA sequence (GBSx1517) was identified in *S.agalactiae* <SEQ ID 4395> which encodes the amino acid sequence <SEQ ID 4396>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 55
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2837(Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15     >GP:BAB04556 GB:AP001510 unknown conserved protein [Bacillus halodurans]
      Identities = 56/185 (30%), Positives = 86/185 (46%), Gaps = 4/185 (2%)

      Query: 7 MDIWTNLGRFAFIETEHVNLRPVAYTDREAFWRIASKRTNLQFT-FPVQTSKKESDFLLV 65
      M+I G +ETE + LR D A + AS +++ + S K+S+ L
      Sbjct: 1 MEIBDIYGDLPFTLETERLRLRFYKDDAAAIYDYASNEQVTKYVLWETHQSIKDSEAFLA 60

20     Query: 66 HSPFMK---EPLGVWAIEDKVSHKMGVIRFENIDLSKKTAEIGYFLKESSWGQGIMTECL 122
      + K + + WAIE K + +M G + F KTAE+GY L E WGQGIMTE +
      Sbjct: 61 FALNKYDEKDVSPWAIELKRNERMIGTVDFVWVKPKDKTAEELGYVLSEPYWGQGIMTEAV 120

      Query: 123 KTLSTFFAFREFGMDKLIIVTHKENIASQKVALKAHFKQSRSFKGSDDRYTRRIRDYIEFQL 182
      L F F +++++ ENI+S +V KA + + + RD+ + +
25     Sbjct: 121 NALVEFGFNNMELERIQAKCFAENISSARVMEKAGLIVEGTHRRRAIVKGAHRDFKVIYAI 180

      Query: 183 TRGDY 187
      R DY
30     Sbjct: 181 IREDY 185

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 667> which encodes the amino acid sequence <SEQ ID 668>. Analysis of this protein sequence reveals the following:

```

35     Possible site: 52
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1096(Affirmative) < succ>
40     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 94/177 (53%), Positives = 117/177 (65%)

45     Query: 7 MDIWTNLGRFAFIETEHVNLRPVAYTDREAFWRIASKRTNLQFI FFPVQTSKKESDFLLVH 66
      MDIWT L FAF ET V LRP Y D F+ + + NL ++FP Q +K SD+LLVH
      Sbjct: 1 MDIWTKLAVFAFFETPKVILRPFRYEDHWDFYSMVNDTKNLYYVFPEQKTKAASDYLLVH 60

      Query: 67 SFMKEPLGVWAIEDKVSHKMGVIRFENIDLSKKTAEIGYFLKESSWGQGIMTECLKTL 126
      SF+K PLG WAIEDK +H++ G IR E+ D + A+IGYFL + WGQGIMTE + L
50     Sbjct: 61 SPIKFPLGQWAIEDKATHQVIGSIRIEHYDAKTRCADIGYFLNYAFWGQGIMTEVVIKLV 120

      Query: 127 FFAFREFGMDKLIIVTHKENIASQKVALKAHFKQSRSFKGSDDRYTRRIRDYIEFQLT 183
      + +F EFG+ L I+TH EN ASQKVA KA F+ FKGS DR T +I Y +QLT
55     Sbjct: 121 YLSFHEFGLKTLRIITHLENKASQKVAKAGFQLKTCFKGS DRNTHKICIIYKMYQLT 177

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1580-

Example 1432

A DNA sequence (GBSx1518) was identified in *S.agalactiae* <SEQ ID 4397> which encodes the amino acid sequence <SEQ ID 4398>. This protein is predicted to be UDP-N-acetylglucosamine-1-carboxyvinyl transferase (murA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 61
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -5.63    Transmembrane    25 - 41 ( 24 - 42)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.3251(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAF86297 GB:AF072894 UDP-N-acetylglucosamine-1-carboxyvinyl
      transferase [Listeria monocytogenes]
      Identities = 240/412 (58%), Positives = 303/412 (73%), Gaps = 2/412 (0%)

20  Query: 3   KIIINGGKQLTGEVAVSGAKNSVVALIPATILADDVVLDGVPASDVSIVDIMETMGA 62
      K+II GGG+L G + V GAKNS VALIPA IIA+ VVL+G+P ISDV +L +I+E +G
      Sbjct: 20 KLIIRGGKKLAGTLQVDGAKNSAVALIPAAIILAESEVVLEGLPDISDVHTLYNILEELGG 79

      Query: 63 KIKRYGETLEIDPCGVKIDIPMPYKINSLRASYFYFGSLLGRYGQATLGLPGGCDLGP RP 122
      ++ +T IDP + +P+P G + LRASY G++LGR+ +A +GLPGGC LGPRP
25  Sbjct: 80 TVRYDNKTAVIDPTDMISMPLPSGNVKKLRASYLLMGAMLGRFKKAVIGLPGGCYLGPRP 139

      Query: 123 IDLHLKAFEAMGASVSYEGDSMLRATNGKPLQGANYMDTVSVGATINTIIAAKANGRT 182
      ID H+K FEA+GA V+ E ++ L + L+GA IY+D VSVGATIN ++AA +A G+T
30  Sbjct: 140 IDQHIGKFEALGAKVTNEQGAIVLRAD--ELKGARIYLDVSVGATINIMLAIVRAKGKT 197

      Query: 183 VIENAAAREPEIIDVATLLNMGAHIRGAGTDVITIEGVKSLHGTRHQVIPDRIEAGTYIA 242
      VIENAA+EPEIIDVATLL NMGA I+GAGTD I I GV+ LHG H +IPDRIEAGT++
      Sbjct: 198 VIENAAKEPEIIDVATLLNMGAIIKGAGTDITRITGVEHLHGCHHTIIPDRIEAGTFMV 257

35  Query: 243 MAAAIGRGIKVTNVLYEHLESFIAKLDEMGMVRMTVEEDSIFVEEQERLKAVSIKTSPPYPG 302
      +AAA G+G+++ NV+ HLE IAKL EMGV M +EED+IFV E E++K V IKT YPG
      Sbjct: 258 LAAASGKGVRIENVIPHTLEGIIAKLTEMGVPMDEDAIFVGEVEKIKKVDIKTYAYPG 317

40  Query: 303 FATDLQQPLTPLLLTAEGNGSLDDTIYEKRVNHVPELARMGANISTLGGKIVYSGPNQLS 362
      F TDLQQPLT LL AEG+ + DTIY R H+ E+ RMG G V +GP QL
      Sbjct: 318 FPTDLQQPLTALLTRAEGSSVITDTIYPSRFKHIAETERMGGKFKLEGRSAVINGPVQLQ 377

      Query: 363 GAPVKATDLRAGAALVIAGLMAEGRTEITNIEFILRGYSNIEKLITSLGADI 414
      G+ V ATDLRAGAALVIA L+A+G TEI +E I RGYS IIEKL+++GA+I
45  Sbjct: 378 GSKVTATDLRAGAALVIAALLADGETEIHGVEHIERGYSKIIEKLISAIGANI 429

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4399> which encodes the amino acid sequence <SEQ ID 4400>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 21
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -8.70    Transmembrane    25 - 41 ( 23 - 45)

55  ----- Final Results -----
      bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

60  >GP:AAF86297 GB:AF072894 UDP-N-acetylglucosamine-1-carboxyvinyl
      transferase [Listeria monocytogenes]
      Identities = 244/412 (59%), Positives = 302/412 (73%), Gaps = 2/412 (0%)

```

-1581-

Query: 3 KIIINGGKALSGEVAVSGAKNSVVALIPAILLADDIVLDGVPASDVDSLIEIMELMGA 62

5 Sbjet: 20 KLIIRGGKLAGTLQVDGAKNSAVALIPAAILAESEVVLGLPDISDVHTLYNILEELGG 79

Query: 63 TVNYHGDLEIDPRGVQDIPMPYKINSLRASYFYFGSLLGRFGQAVVGLPGGCDLGPRP 122

10 Sbjet: 80 TVRYDNKTAVIDPTDMLSMPLPSGNVKKLRASYLGMAGLGRFKKAVIGLPGGCYLGP 139

Query: 123 IDLHLKAFEAMGVEVSIEGENMNLSTNGQKIHGAHIYMDTVSVGATINTMVAATKAQGT 182

15 Sbjet: 140 IDQHKGFEALGAKVTNEQGAITYLRAD--ELKGARIYLDVSVGATINIMLA AAVRAKGKT 197

Query: 183 VIENAAAREPEIIDVATLLNNMGAIHAGTDDIITIQQVQKLHGTRHQVIPDRIEAGTYIA 242

20 Sbjet: 198 VIENAAAREPEIIDVATLLNNMGAIHAGTDDIIRITGVHHLHGCHHTIIPDRIEAGTFMV 257

Query: 243 LAAAIGKGVKITNVLYEHLESFIAKLEEMGVRMTVEEDAIFVEKQESLKAITIKTSPYPG 302

25 Sbjet: 258 LAAASGKGVRIENVIPHTLEGIIAKLTEMGVPMDEEDAIFVGEVEKIKKVDIKTYAYPG 317

Query: 303 FATDLQQPLTPLLKADGRGTIIDTIYEKRINHVPPELMMRGADISVIGGQIVYQGPSRLT 362

30 Sbjet: 318 FPTDLQQPLTALLTRAEGSSVITDTIYPSRFKHIAETERMGGKFKLEGRSAVINGPVQLQ 377

Query: 363 GAQVKATDLRAGAALVITAGLIAEGKTEITNIEFILRGYASIIAKLTALGADI 414

35 Sbjet: 378 GSKVTATDLRAGAALVIAALLADGETEIHGVEHIERGYSKIEKLSAIGANI 429

An alignment of the GAS and GBS proteins is shown below.

Identities = 344/419 (82%), Positives = 394/419 (93%)

35 Query: 1 MRKIIINGGKQLTGEVAVSGAKNSVVALIPATILADDVVLDGVPASDVDSLVDIMETM 60

Sbjet: 1 MRKIIINGGK L+GEVAVSGAKNSVVALIPA ILADD+V+LDGVPASDVDSL++IME M

Query: 61 GAKIKRYGETLEIDPCGVKDIPMPYKINSLRASYFYFGSLLGRYQATLGLPGGCDLGP 120

40 Sbjet: 61 GATVNYHGDLEIDPRGVQDIPMPYKINSLRASYFYFGSLLGRFGQAVVGLPGGCDLGP 120

Query: 121 RPIDLHLKAFEAMGASVSIEGDSMRLATNGKPLQGANIYMDTVSVGATINTIIAAKANG 180

45 Sbjet: 121 RPIDLHLKAFEAMG VSYEG++M L+TNG+ + GA+IYMDTVSVGATINT++AA KA G

Query: 181 RTVIENAAAREPEIIDVATLLNNMGAIHAGTDDVITIEGVKSLHGTRHQVIPDRIEAGTY 240

50 Sbjet: 181 KTVIENAAAREPEIIDVATLLNNMGAIHAGTDDIITIQQVQKLHGTRHQVIPDRIEAGTY 240

Query: 241 IAAAAIGRGIKVTNVLYEHLESFIAKLEEMGVRMTVEEDSIFVEEQERLKAVSIKTSFY 300

55 Sbjet: 241 IALAAAIGKGVKITNVLYEHLESFIAKLEEMGVRMTVEEDAIFVEKQESLKAITIKTSFY 300

Query: 301 PGFATDLQQPLTPLLTAEGNGSLDDTIYEKRVNHVPELARMGANISTLGGKIVYSGPNQ 360

60 Sbjet: 301 PGFATDLQQPLTPLL A+G G+++DTIYEKR+NHVPEL RMGA+IS +GG+IVY GP++

Query: 361 LSGAPVKATDLRAGAALVIAGLMAEGRTEITNIEFILRGYSNIEKLTSLGADIQLVEE 419

60 Sbjet: 361 LTGAQVKATDLRAGAALVITAGLIAEGKTEITNIEFILRGYASIIAKLTALGADIQLIED 419

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1582-

Example 1433

A DNA sequence (GBSx1519) was identified in *S.agalactiae* <SEQ ID 4401> which encodes the amino acid sequence <SEQ ID 4402>. This protein is predicted to be thiamine phosphate pyrophosphorylase (thiE). Analysis of this protein sequence reveals the following:

```

5      Possible site: 55
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.0422(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:AAF25544 GB:AF109218 ThiE [Staphylococcus carnosus]
      Identities = 98/200 (49%), Positives = 140/200 (70%), Gaps = 1/200 (0%)

      Query: 5   LKLYFVCGTVDCSR-KNILTVEEALQAGITLQFREKGFALQGKEKIAMAKQLQILCK 63
      L +YF+CGT D + I V++EAL+ GITL+QFREKG A G++K+A+AK+LQ LCK
20      Sbjct: 7   LNVYFICGTQDIPEGRTIQEVLKEALEGGITLYQFREKNGAKTGQDKVALAKELQALCK 66

      Query: 64   QYQVPFIIDDDIDLVELIDADGLHIGQNDLPVDEARRRLPDKIIGLSVSTMDEYQKSQLS 123
      Y VPF++DD+ L E IDADG+H+GQ+D VD+ R KIIGLS+ ++E S L+
25      Sbjct: 67   SYNVPFIVNDDVALAEEIDADGIHVGDDEAVDDFNNRFEGKIIGLSIGNLEELNASDLT 126

      Query: 124  VVDYIGIGPFNPTQSKADAKPAVGNRTTKAVREINQDPIVAIGGITSDFVHDIIESGAD 183
      VDYIG+GP T SK DA VG + + +R+ D+PIVAIGGI+ D V ++ ++ AD
30      Sbjct: 127 VVDYIGVGPIFATPSKDDASEVGPVKMIETLRKEVGDLPIVAIGGISLDNVQEVAKTSAD 186

      Query: 184  GIAVISAISKANHIVDATRQ 203
      G++VISAI+++ H+ + +
35      Sbjct: 187 GVSVISAIARSPHVTETVHK 206

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1434

A DNA sequence (GBSx1520) was identified in *S.agalactiae* <SEQ ID 4403> which encodes the amino acid sequence <SEQ ID 4404>. This protein is predicted to be hydroxyethylthiazole kinase (b2104). Analysis of this protein sequence reveals the following:

```

40      Possible site: 54
      >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL Likelihood = -4.94 Transmembrane 198 - 214 ( 194 - 217)

      ----- Final Results -----
45      bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8805> which encodes amino acid sequence <SEQ ID 8806> was also identified. Analysis of this protein sequence reveals the following:

```

50      Lipop: Possible site: -1 Crend: 7
      McG: Discrim Score: -2.93
      GvH: Signal Score (-7.5): 1.61
      Possible site: 39
55      >>> Seems to have no N-terminal signal sequence

```


-1583-

ALOM program count: 1 value: -4.94 threshold: 0.0
 INTEGRAL Likelihood = -4.94 Transmembrane 183 - 199 (179 - 202)
 PERIPHERAL Likelihood = 2.49 151
 modified ALOM score: 1.49

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF25543 GB:AF109218 ThiM [Staphylococcus carnosus]
 Identities = 114/253 (45%), Positives = 160/253 (63%), Gaps = 1/253 (0%)

Query: 18 LEQLKEVNPLTICITNNVKNFTANGLLALGASPAMSECIEDLELLKVADALLINIGTL 77
 L+Q++ +PL IC TN+VVKNFTANGLL+LGASP MSE ++ ED VA ++LINIGTL
 Sbjet: 5 LDQIRTEHPLVICYTNDVVKNFTANGLLSLGASPTMSEAPQEAEDFYFVAGSVLINIGTL 64

Query: 78 TKESWQLYQEAIKIANKNQVPVVLDPVAGASRFRLEVSLDLLKNYSISLLTNGNSEIAA 137
 TK E KIAN+ + P+V DPVA GAS++R + LK +++ GN SEI A
 Sbjet: 65 TKHHEHAMLENAKIANETETPLVDFPVAVGASKYRKDFCKYFLKKIKPTVIKGNASEILA 124

Query: 138 LIGEKQASKGADGGKVADLESIAVKANQVDFVPVVVTGETDAIAVRGEVRLQLNGSPLMP 197
 LI + KG D D+ IA KA + + +++TGETD I +V L NGS +
 Sbjet: 125 LIDDTATMKGTSADNLDVVDIAEKAYKEYQTAILTGETDVIVQDNKVVKLSNGSHFLA 184

Query: 198 LVTGTGCLLGAVLAFIGSSDRSDDIACLTTEAMTVYNVAGEIAEKVAKGKGVGSFQVAF 257
 +TG GCLLGAV+ AF+ + + L EA++VYN+A E AE+++ KG G+F F+
 Sbjet: 185 KITGAGCLLGAVVGAFL-FRNTHPSIETLIEAVSVYNIAAERAEQLSDSKGPGTFLTQFI 243

Query: 258 DALSQMKSEMIMD 270
 DAL ++ S+ + +
 Sbjet: 244 DALYRIDSDAVAE 256

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8806 (GBS398) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 6; MW 31.8kDa).

The GBS398-His fusion product was purified (Figure 214, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 314), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1435

A DNA sequence (GBSx1521) was identified in *S.agalactiae* <SEQ ID 4405> which encodes the amino acid sequence <SEQ ID 4406>. This protein is predicted to be ThiD (thiD). Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1584-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF25542 GB:AF109218 ThiD [Staphylococcus carnosus]
Identities = 139/258 (53%), Positives = 186/258 (71%), Gaps = 4/258 (1%)

5 Query: 8 LTIAGTDPGAGIMADLKTQARRTYGMAVVTSVVAQNTCGVRGVQHIETAIIDQQLAC 67
      LTIAGTDP+GGAG+MADLK+F A YGMA +TS+VAQNT GV+ + +++ + +QL
Sbjct: 8 LTIAGTDPGAGVMADLKSFHACGVYGMAAITSIVAQNTKGVQHIHNLDTWLKEQLDS 67

10 Query: 68 VYDDIKPKAVKTGMLAERETISLVASYLKYPQ-PYVLPVMVATSGHRLIDSDAVEALK 126
      ++DD P+A+KTGM+A +E + L+ SYL+KYP PYV+DPVM+A SG L+D AL+
Sbjct: 68 IFDDELPPQAIKTGMIATKEMMELIRSYLEKYPDIPYVIDPVMMLAKSGDSLMDAGKHALQ 127

15 Query: 127 EDLLPLATIITPNLPEAEVLVGVDLSDEVSIKAGYDIQKQYSVRNVLIKGGHLD--GLA 184
      E LLPLA + TPNLPEAE +VG+ L E +I KAG + + V+IKGGH++ +A
Sbjct: 128 BILLPLADVATPNLPEAEIVGFKLDTEEAIKKAGDIFINEIGSKGVVIKGGHIEDKNIA 187

20 Query: 185 KDYLEKEKEGLITLSNQRIINTIHTGTGCTFAAVVAELAKGQSILNAVSTAKSFITSAI 244
      KDYLEF K+GL ++R +T HTHGTGCTF+AV+ AELAKG++I AV AK FI +I
Sbjct: 188 KDYLEF-TKDGLEVFESERYDTKHTHTGTGCTFSAVITAEELAKGKTIYEA VKKAKDFIALSI 246

25 Query: 245 ETAPELGLGNGP VNHNTSY 262
      + PE+G G GPVNH +Y
Sbjct: 247 KYTPEIGQGRGPVNH FAY 264
```

25 There is also homology to SEQ ID 4408.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1436

30 A DNA sequence (GBSx1522) was identified in *S.agalactiae* <SEQ ID 4409> which encodes the amino acid sequence <SEQ ID 4410>. This protein is predicted to be TenA (tenA). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
```

```
35 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2242(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

40 The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF25541 GB:AF109218 TenA [Staphylococcus carnosus]
Identities = 78/213 (36%), Positives = 127/213 (59%), Gaps = 6/213 (2%)

45 Query: 14 IQSIYQDPFIQGIKGRLDHVDVICHYLDQADNIYLGKFADIIYALCLAKSDNLRDKQFFLEQ 73
      I IYQD FIQ ++KG + + + YL+AD YL +FA+IYAL + +L +F ++Q
Sbjct: 15 IDEIYQDHFIQELLKGDIKKEALRQYL RADASYLREFANIIYALLIPIMPDLSESVRFLVDQ 74

50 Query: 74 IDFTLNRELADGEGPHQALAAAYTNRSYQDIIEKGVWYPSADHYIKHMYFHFY-ENGIAGA 132
      I F +N E+ H+ +A Y +Y +I+K VW PS DHYIKHMY++ Y A A
Sbjct: 75 IQFIVNGEVE----AHEYMADYIGENYNEIVQKKVWPPSGDHYIKHMYYNVYAHENAAYA 130

55 Query: 133 LAAMSPCPWIYHQLAKKIIENQFLNGNPFNNWITFYANDTVEELMENYFRMMDYYAQN 192
      +AAM+PCP++Y +AK+ +++ + W FY N ++ L+E +M+ N+
Sbjct: 131 IAAMAPCPYVYAMIAKRAMKDPNLNKSSILAKWFEFY-NTEMPLIEVLDDLMNQLTANM 189

Query: 193 SKEKQADLVDAFVKSCQHERRFFQMAINQEKWE 225
      S+ ++ ++ + +++S HE FF MA EKW+
Sbjct: 190 SETEKNEVRENYLQSTVHELNFNFMAYTSEKWQ 222
```

-1585-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1437

- 5 A DNA sequence (GBSx1523) was identified in *S.agalactiae* <SEQ ID 4411> which encodes the amino acid sequence <SEQ ID 4412>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have a cleavable N-term signal seq.

10 INTEGRAL Likelihood = -7.06 Transmembrane 43 - 59 (36 - 63)
 INTEGRAL Likelihood = -2.55 Transmembrane 92 - 108 (92 - 112)
 INTEGRAL Likelihood = -1.49 Transmembrane 135 - 151 (135 - 151)
 INTEGRAL Likelihood = -1.06 Transmembrane 69 - 85 (69 - 85)
 INTEGRAL Likelihood = -0.22 Transmembrane 216 - 232 (216 - 232)

15 ----- Final Results -----

bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA91230 GB:Z56283 orf2 [Lactobacillus helveticus]

Identities = 46/215 (21%), Positives = 96/215 (44%), Gaps = 3/215 (1%)

25 Query: 21 AITFLCLLIPTFSFSFTLRRLRTSLFLIIVVTLQCFVKVSLKTWAKVNLISFVMGLSLFL 80
 ++ F+ I + S L T+L+ + + ++ +K + + F+ ++F
 Sbjet: 4 SLKFILAFIISLEISLKASLTNTNLIVIAFALIYLLVTRIKIKELILLIIVPFIASFTIFA 63

30 Query: 81 GTYFWGKLPHQFVLASLVACRPLIFMNVGLLFHASHSNYDFIESLYQTFKVPSPHFAYGIF 140
 +++ P + +L + R ++ + + DF SL Q +PS FAYG+
 Sbjet: 64 TLFWFSPTPDAYYAWNLS-STRVYVYTLTIACVTRNTTATDFARSLEQNLHLPSKFAYGV 122

35 Query: 141 AVFNLLPLIKLQYQRNRLAFRLKNQVTWALSPRLILSVLLKTIYWVEQLELAMLKGF 200
 A N++P +K ++ R + ++ SP L +L + + L M S G+
 Sbjet: 123 AAINIIPRMKTAVKQIRTSAMMRGMYLSFWSPVLYFKAILVALNSADNLAQGMESHG 182

Query: 201 GKERTHASTYPVRFDRDYSL-LGMSILLSIGM-IFK 233
 G++R P+ +D+ + + IL++I + IFK
 Sbjet: 183 GQKRATIVAIPLTKKDWLIFFTLILVNISLFIK 217

- 40 No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8807> and protein <SEQ ID 8808> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0

McG: Discrim Score: 4.50

45 GvH: Signal Score (-7.5): -0.2

Possible site: 35

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 5 value: -7.06 threshold: 0.0

50 INTEGRAL Likelihood = -7.06 Transmembrane 43 - 59 (36 - 63)
 INTEGRAL Likelihood = -2.55 Transmembrane 92 - 108 (92 - 112)
 INTEGRAL Likelihood = -1.49 Transmembrane 135 - 151 (135 - 151)
 INTEGRAL Likelihood = -1.06 Transmembrane 69 - 85 (69 - 85)
 INTEGRAL Likelihood = -0.22 Transmembrane 216 - 232 (216 - 232)
 PERIPHERAL Likelihood = 2.65 170
 55 modified ALOM score: 1.91

*** Reasoning Step: 3

-1586-

----- Final Results -----

bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1438

A DNA sequence (GBSx1524) was identified in *S.agalactiae* <SEQ ID 4413> which encodes the amino acid sequence <SEQ ID 4414>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15

bacterial cytoplasm --- Certainty=0.3007(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20

>GP:CAA91229 GB:Z56283 orf1 [Lactobacillus helveticus]

Identities = 123/424 (29%), Positives = 200/424 (47%), Gaps = 48/424 (11%)

25

Query: 17 LFDEVTFSLNPGERILISGYSGCGKSTLALLLSGL--KESGK--GQVLLNGSLIEPSDVG 72
 L +++ ++ PG +LI G +GCGKSTL +++GL K +GK G++ L+G
 Sbjct: 12 LINQNLNMNIAPGFNLLI-GPTGCGKSTLLKIIAGLYPKYAGKLTGKIDLHGQ-----KAA 65

30

Query: 73 FLFQNPDLQFCMDTVAHELYFILENLQIEPEQMQRSEFVLAQVGLKGFQNRLIYTLISQG 132
 +FQN QF M T E+ F LENLQI+ + + + + ++ I TLS G
 Sbjct: 66 MMFQNAEQFTMTTPREEIIFALENLQIKAKDYDLHIKKAFTKIDLLDQKINTLSGG 125

35

Query: 133 EKQRLALATIFLKSPKLIILDEAFANLDQESASQLLQVLNLYQANNQSMILVIDHLITYY 192
 ++Q +ALA + + +LDE FA+ D + L++ + + ++ +I+ DH++ Y
 Sbjct: 126 QQQHVALAVLIAMDVDVFLLEPFASCDPNTRHFLIEKLASLAETGRT-IILSDHVLDDY 184

40

Query: 193 QDIMDHYFWLEKRLTRVNFYMLNRLNVFELEKKSHN-----TGDKLLSLKDFQVK- 243
 + I DH + E + + N+L F+ K+ H TG + + Q+K
 Sbjct: 185 EKICDHLVQFEGKTVKELSANENKLI--FKQNKQFHEQSYFALPTGTPVFELNKTQIKQ 242

45

Query: 244 ----LSKNKFISYLDLDFLASGERLCLDGPVGKSSLFMGLLGLYRTKGK-----KQ 291
 L +NK Y G+ + G +GVGK+SLF + + KG +
 Sbjct: 243 NRLLKQNKLIY-----GKTTTLITGSGNGVGKTSLFKAMTKMIPYKGNFTYLDNEISK 295

50

Query: 292 FTHRQKIP-ISFLFQNPDLQFIFSTVYDEIFQVCKDSN-----KARDILETINLWDKKQ 344
 +RK + I+ FQ DQF+ TV DEI KD N K + LE + L
 Sbjct: 296 IKYRKYSQIAQFFQKASDQFLTIVTKDEIELSKDRNNFFTDKIDEWLEKLQKQHL 355

Query: 345 FSPFQLSQGQORRLAIGSILASDSKLLLLDEPTYQDAYHANMITTLLSYCHKNHCGVI 404
 + LS GQQ++L I +L + +LL+DEP G D +++ L+ K +
 Sbjct: 356 QVVYSLSGGQKKLQILLMLMTKHNVLIDEPLSGLDHESVDLVLQLMQECQEKLQQTFL 415

Query: 405 FTSH 408

SH

Sbjct: 416 IISH 419

55

Identities = 44/185 (23%), Positives = 83/185 (44%), Gaps = 24/185 (12%)

Query: 28 GERILISGYSGCGKSTLALLLSGLKESGKGQVLLNGSLIEP-----SDVGFLFQNPDLQ 81
 G+ LI+G +G GK++L ++ + L+ + + S + FQ Q
 Sbjct: 256 GKTTTLITGSGNGVGKTSLFKAMTKMIPYKGNFTYLDNEISKIKYRKYSQIAQFFQKASDQ 315

60

Query: 82 FCMDTVAHELYFILENLQIEPEQMQRSEFV-----LAQVGLKGFQNRLIYTLISQGE 133
 F TV E+ +DR+ F L ++ LK ++++Y+LS G+

-1587-

Sbjct: 316 FLTIVTKDRIEL-----SKKDRNNFFTDKIDEWLEKQLKQHLQVVSLSGGQ 365

Query: 134 KQRLALATIFLKSPKLIILDEAFANLDQESASQLLQVLNYQANNQSMILVIDHLITYYQ 193

5 Sbjct: 366 QKKLQILLMLMTKHNVLIDEPLSGLDHESVDLVLQLMQECQEKIQQTFLLIISHQIDALA 425

Query: 194 DIMDH 198

D D+

10 Sbjct: 426 DFCDY 430

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4415> which encodes the amino acid sequence <SEQ ID 4416>. Analysis of this protein sequence reveals the following:

Possible site: 30

15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3093(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 120/455 (26%), Positives = 203/455 (44%), Gaps = 47/455 (10%)

25 Query: 1 MLSEVKIACHTGDSHYLFDEV-TFSLNPGERILISGYSGCGKSTLALLSGLKE---SGK 56
M+S E+L T+ D ++ T + G+ I++ G SG GKST LL+G+ +GK

Sbjct: 21 MISAEQLVFTYHDQKNPACQISTCQIASGQFIVLCGPGSGSGKSTFLKLLNGIIPDYAGK 80

30 Query: 57 GQVLLNGSLIEPS-----DVGFLFQNPDLQFCMDTVAHELYFILENLQIEPEQM 107
+ L+ + + V +FQNP QF V HEL F EN ++ + +

Sbjct: 81 YEGRLDVADCCQAGRDSVETFSRSVASVFQNPASQFFYREVQHELVPFCENQGLDAKVMK 140

35 Query: 108 RSEFVLAQVGLKGFQNRILIYTLISQGEKQRLALATIFLKSPKLIILDEAFANLDQESASQL 167
R + N+ ++ LS G+KQR+A+AT ++ +++ DE ANLD + +

Sbjct: 141 RLWTLAEDFAFAELLNKDMFGLSGGQKQKRVAIATAIMQGTNIMLFDEPTANLDSAGIAAV 200

Query: 168 LQLVLNYQANNQSMILVIDHLITYYQDIMDHYFW-----LEKRLTRVNF-----DY 213
+ +A ++ +IV +H + Y D+ D++F+ L +LT N D

Sbjct: 201 KAYLTQLKAAGKT-IIVAERHLHYLMDLADNFFYFKNGRLTDKLITONLLALTDEQRQDM 259

40 Query: 214 MLNRLNVFELE-----KKSHTNGDKLLSIKDFQVKLSKNKFISYLDLDFLASGERLCLD 266
L RL++ +L+ .. + H D L I+ V+ A G +

Sbjct: 260 GLRRDLSDLKPVLAGKIESQHYRPDDSLCIEHLTVRAGSKILRCIEQLSFAVGSISGIT 319

45 Query: 267 GPSGVGKSSLFMGLLGLYRTKGGKQFTHRKQIPISFLFQNPDLQFIFSTVYDEIF--QVC 324
G +G+GKS L + G+ KK + IP+S + + V ++F V

Sbjct: 320 GSNGLGKSQLVYYIAGI--LDDKKATIKFQGIPLSAKQRLSKTSIVLQEVSLQLFAESVS 377

50 Query: 325 KDSN-----KARDILETINLWDKKQFSPFQLSQGQRRRLAIGSILASDSKLLLLDEPT 377
K+ N + +++E ++L + P LS G+Q+R+ I + L +D +L+ DEP+

Sbjct: 378 KEVNLGHERHPRTEVIERLSLITLLERHPASLSGGEQQRVMAASLLADKIDILIFDEPS 437

Query: 378 YGQDAYHANMITLLLSYCHKNHCGVIFTSHDPHL 412
G D + LL+ H VI SHD L

55 Sbjct: 438 SGLDLLQMKALANLLMQ-LKTQHKVVILISHDEEL 471

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1439

60 A DNA sequence (GBSx1525) was identified in *S.agalactiae* <SEQ ID 4417> which encodes the amino acid sequence <SEQ ID 4418>. Analysis of this protein sequence reveals the following:

-1588-

Possible site: 42

>>> Seems to have an uncleavable N-term signal seq

5 INTEGRAL Likelihood = -11.62 Transmembrane 8 - 24 (1 - 30)
 INTEGRAL Likelihood = -8.17 Transmembrane 145 - 161 (143 - 163)
 INTEGRAL Likelihood = -6.32 Transmembrane 66 - 82 (62 - 84)
 INTEGRAL Likelihood = -3.77 Transmembrane 112 - 128 (111 - 132)
 INTEGRAL Likelihood = -2.66 Transmembrane 43 - 59 (43 - 59)

----- Final Results -----

10 bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAB13180 GB:Z99110 ykoE [Bacillus subtilis]
 Identities = 68/177 (38%), Positives = 117/177 (65%), Gaps = 1/177 (0%)

 Query: 5 LKDVLLIALLAIVLGVVYFGAGYISNAFVFPVGPITAEVYIGIWFVAGPMALYILRKPGT 64
 +K++++++V VVY + N GPIA+E IYGIWF+ +A Y++RKPG
 20 Sbjct: 6 VKEIVIMSVISIVFAVVYLLFTHFGNVLAGMFGPIAYEPIYGIWFIVSVIAAYMIRKPGA 65

 Query: 65 AIVAEIIAALIEVLIGSIYGPSVLVIGTQLGLGSELGFTLFRYHNYKLPAFILSAILTSI 124
 A+V+E++AAL+E L+G+ GP V+VIG +QGLG+E F R+ Y LP +L+ + +S+
 25 Sbjct: 66 ALVSEIIAALVECLLGNPSGPMVIVIGVQGLGAEAVFLATRWKAYSLPVLMLAGMGSSV 125

 Query: 125 FSAWFSFYANGLSAFSAFSYNILMLIVRTVS-SIIFFLLTKNICDQLHRSGVLNAYGI 180
 SF + + +G +A+S Y ++ML++R +S +++ LL K + L +GVLN +
 30 Sbjct: 126 ASFIYDLFVSGYAAAYSPGYLLIMLVIRLISGALLAGLLGRAVSGSLAYTGVLNGMAL 182

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1440

35 A DNA sequence (GBSx1526) was identified in *S.agalactiae* <SEQ ID 4419> which encodes the amino acid sequence <SEQ ID 4420>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

40 INTEGRAL Likelihood = -6.69 Transmembrane 65 - 81 (53 - 95)
 INTEGRAL Likelihood = -6.37 Transmembrane 34 - 50 (31 - 54)
 INTEGRAL Likelihood = -6.10 Transmembrane 176 - 192 (169 - 195)
 INTEGRAL Likelihood = -3.66 Transmembrane 130 - 146 (130 - 151)
 INTEGRAL Likelihood = -1.97 Transmembrane 3 - 19 (3 - 19)
 INTEGRAL Likelihood = -0.90 Transmembrane 88 - 104 (88 - 104)

45 ----- Final Results -----

 bacterial membrane --- Certainty=0.3675(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 A related GBS nucleic acid sequence <SEQ ID 9757> which encodes amino acid sequence <SEQ ID 9758> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

55 A related GBS gene <SEQ ID 8809> and protein <SEQ ID 8810> were also identified. Analysis of this protein sequence reveals the following:

-1589-

Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: -4.09
 GvH: Signal Score (-7.5): -4.38
 Possible site: 47

5 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 6 value: -6.69 threshold: 0.0
 INTEGRAL Likelihood = -6.69 Transmembrane 65 - 81 (53 - 95)
 INTEGRAL Likelihood = -6.37 Transmembrane 34 - 50 (31 - 54)
 INTEGRAL Likelihood = -6.10 Transmembrane 176 - 192 (169 - 195)
 10 INTEGRAL Likelihood = -3.66 Transmembrane 130 - 146 (130 - 151)
 INTEGRAL Likelihood = -1.97 Transmembrane 3 - 19 (3 - 19)
 INTEGRAL Likelihood = -0.90 Transmembrane 88 - 104 (88 - 104)
 PERIPHERAL Likelihood = 5.30 158
 modified ALOM score: 1.84

15 *** Reasoning Step: 3

----- Final Results -----

20 bacterial membrane --- Certainty=0.3675(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 **Example 1441**

A DNA sequence (GBSx1527) was identified in *S.agalactiae* <SEQ ID 4421> which encodes the amino acid sequence <SEQ ID 4422>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have a cleavable N-term signal seq.

30 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 A related GBS gene <SEQ ID 8811> and protein <SEQ ID 8812> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 2
 McG: Discrim Score: 6.01
 GvH: Signal Score (-7.5): 0.45
 45 Possible site: 23
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 10.66 threshold: 0.0
 PERIPHERAL Likelihood = 10.66 80
 modified ALOM score: -2.63

50 *** Reasoning Step: 3

----- Final Results -----

55 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1590-

SEQ ID 4422 (GBS19) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 4; MW 24kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 9 (lane 6; MW 46.1kDa).

The GST-fusion protein was purified as shown in Figure 190, lane 10.

5 Example 1442

A DNA sequence (GBSx1528) was identified in *S. agalactiae* <SEQ ID 4423> which encodes the amino acid sequence <SEQ ID 4424>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8813> which encodes amino acid sequence <SEQ ID 8814> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1   Crend: 6
SRCFLG: 0
McG: Length of UR: 23
      Peak Value of UR: 2.61
      Net Charge of CR: 3
McG: Discrim Score: 9.08
GvH: Signal Score (-7.5): -0.76
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 23
ALOM program count: 0 value: 5.14 threshold: 0.0
      PERIPHERAL Likelihood = 5.14 365
      modified ALOM score: -1.53

*** Reasoning Step: 3

Rule gp01

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA34476 GB:X16457 precursor polypeptide (AA -26 to 632)
[Staphylococcus aureus]
Identities = 93/372 (25%), Positives = 160/372 (43%), Gaps = 46/372 (12%)

Query: 9  MKKQFLKSAAILSLAVTAVSTSQPVGAIVGKDETKLRQQLGYIDSKKSGKKIDERWGEKI 68
      MKKQ + A L++A + + AIV KD +K + + K G + + + KI
Sbjct: 1  MKKQIISLGA-LAVASSLFTWDNKADAIVTKDYSK---ESRVNEKSKKGATVSDYYYWKI 56

Query: 69  YNYLSYELIEANEWINRSEFQEPEYRTILSEFKDKIDSIEYYLINLS----NIAKEDAHQ 124
      + L + A + + ++ +P Y+ ++ + YL+ + K+
Sbjct: 57  IDSLEAQFTGAIDLLENYKYGDPIYKEAKDRMLTRVLGEDQYLLKKKIDYEYLYKKWYKS 116

Query: 125 RNILQSLDKYEKSGIYNLDQGVYNYIYQEISSAKHKFSDGVDKIYRLDSTLFFFSVWYDK 184
      N ++ + K +YNL YN I+ + A ++F+ V +I + L F
Sbjct: 117 SNKNTNMLTFHKYNLYNLTMNEYNDIFNSLKDQAVYQFNKEVKIEHKNVDLKQF----- 170

Query: 185 HLDNNDNYKDNKDFKEYIALLNETTRKARLGYQIVNNHKD-GEHKDEAEI-LDILIRDIT 242

```


-1591-

```

          D  ++K KE  L++EI      Y      KD GEH E   LD+++ D
Sbjct: 171 -----DKDGEDKATKEVYDLVSEIDTIVVTFYYA-----DKDYGEHAKELRAKLDLILGDT 221

Query: 243 FVSKDAPGYKYIPNKRIAAKIIEDLDGIINDFFKNTGDKP-SLEKLKDTEFHKKYLNST 301
          K      I N+RI  ++I+DL+ II+DFF T +++P S+ K   T+ + K  +
Sbjct: 222 NPHK-----ITNERIKKEMIDDLNSIIDFFMETKQNRPNSTIKYDPTKHNFKSESEN 274

Query: 302 EPYSIETNLPSNYKELKEKQIKKLEYGYK-KSSKIY--TSAHYALYSEEIDAAKELLQKV 358
          +P   N      +E K K +K+ +  +K K+ K Y T      + EE   + L KV
Sbjct: 275 KP-----NFDKLVEETK-KAVKEADESWKNKTVKYEETVTKSPVVKEEKKVEEPQLPKV 328

Query: 359 KIAKDNYNEIKS 370
          N E+K+
Sbjct: 329 ----GNQQEVKT 336

```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8814 (GBS119) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 2; MW 84.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 5; 2 bands).

The GBS119-GST fusion product was purified (Figure 109A; see also Figure 201, lane 6) and used to immunise mice (lane 1+2+3 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS (Figure 109B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1443

A DNA sequence (GBSx1529) was identified in *S.agalactiae* <SEQ ID 4425> which encodes the amino acid sequence <SEQ ID 4426>. This protein is predicted to be s-adenosylmethionine synthetase (metK). Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3609(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:EAB07019 GB:AP001518 S-adenosylmethionine synthetase [Bacillus halodurans]
Identities = 266/390 (68%), Positives = 324/390 (82%), Gaps = 1/390 (0%)

Query: 4   RKLFTSESVSEGHDPKIDQISDAILDAILEQDPDAHVAETA VYTGSVHVFG EISTTAY 63
          R+LFTSESV+EGHPDKI DQISD+ILD IL++DP+A VA ET+V TG V V GEI+T+ Y
Sbjct: 7   RRLFTSESVTEGHDPKICDQISDSILDEILKEDPNARVACETSVTTGLVLVAGEIITSTY 66

Query: 64  VDINRVVRNTIAEIGYDKAEYGFSAESVGVHPSLVEQSPDIAQGVNEALEVR-GSLEQDP 122
          VDI +VVR+TI IGY +A+YGF +E+ V S+ EQSPDIAQGVN+ALE R G +
Sbjct: 67  VDIPKVVVRTIRNIGYTRAKYGFDS ETCVLTSLIDEQSPDIAQGVNQALEAREGQMTDAE 126

Query: 123 LDLIGAGDQGLMFGFAVD ETP ELMPLPISLAHQLVKKLTDLRKSGELTYLRPD AKSQVTV 182
          ++ IGAGDQGLMFG+A +ETPELMPLPISL+H+L ++L++ RK L YLRPD K+QVTV
Sbjct: 127 IEAIGAGDQGLMFGYANNETPELMPLPISLSHKLARRLSEARKGEILPYLRPDGKTQVTV 186

Query: 183 EYDENDQPIRVDAVVISTQHDPNVINDQLHKDVIEKVINEVIPSHYLD DQTKFFINPTGR 242

```

-1592-

EYDENDQ +R+D +VISTQH P VT +Q+ D+ + VI V+P +D++TK+FINPTGR
 Sbjct: 187 EYDENDQSVRIDTIVISTQHHPVETLEQIESDLKQHVIRSVVPEELIDEETKYFINPTGR 246
 Query: 243 FVIGGPQGD SGLTGRKIIIVDTYGGYSRHGGGAFSGKD ATKVDRSASAYAARYIAKNIVAAD 302
 FVIGGPQGD+CLTGRKIIIVDTYGGY+RHGGGAFSGKD TKVDRS +YAARY+AKNIVAA
 Sbjct: 247 FVIGGPQGDAGLTGRKIIIVDTYGGYARHGGGAFSGKDPTKVDRSGAYAARYVAKNIVAAG 306
 Query: 303 LAKKVEVQLAYAIQVAPVSVRVDTFGTGVIAEADLEAAVRQIFDLRPAGIINMLDLKRP 362
 LA K EVQLAYAIQVAPVSV +DTFGTG ++EA L VR+ FDLRPAGII MLDL+RP
 Sbjct: 307 LADKCEVQLAYAIQVAPVSVISIDTFGTGQVSEARLVELVREHFDLRPAGIIMLDLRRP 366
 Query: 363 IYRQTAAAYGHMGRDIDLPWERVDKVQALK 392
 IY+QTAAAYGH GRD++LPWE+ DK + L+
 Sbjct: 367 IYKQTAAAYGHFGRDVELPWEQTDKAEILR 396

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4427> which encodes the amino acid sequence <SEQ ID 4428>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3389(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 333/395 (84%), Positives = 361/395 (91%), Gaps = 1/395 (0%)
 Query: 1 MSERKLTSESSEVSEGHDPKIDQISDAILDAILEQDPDAHVAEAETAVYTGSVHVFGEIST 60
 MSERKLTSESSEVSEGHDPKIDQISDAILDAIL +DP+AHVAEAET VYTGSVHVFGEIST
 Sbjct: 1 MSERKLTSESSEVSEGHDPKIDQISDAILDAILEQDPDAHVAEAETCVYTGSVHVFGEIST 60
 Query: 61 TAYVDINRVVRNTIAEIGYDKAEYGFSAESVGVHPSLVEQSPDIAQGVNEALEVRGSLEQ 120
 TAY+DINRVVR+TIAEIGY +AEYGFSAESVGVHPSLVEQS DIAQGVNEA E R +
 Sbjct: 61 TAYIDINRVVRDITAEIGYTEAEYGFSAESVGVHPSLVEQSGDIAQGVNEAFESREG-DT 119
 Query: 121 DPLDLIGAGDQGLMFGFAVDETPELMPLPISLAHQLVKKLTDLRKSGELTYLRPDAKSQV 180
 D L IGAGDQGLMFGFA++ETPELMPLPISL+HQLV++L +LRKSGE++YLRPDAKSQV
 Sbjct: 120 DDLSHIGAGDQGLMFGFAINETPELMPLPISLSHQLVRRILAE LRKSGEISYLRPDAKSQV 179
 Query: 181 TVEYDENDQPIRVDAVVISTQHDPNVTNDQLHKDVIEKVINEVIPSHYLDQTKFFINPT 240
 TVEYDE+D+P+RVD VVISTQHDP TNDQ+ +DVIEKVI VIP+ YLDD TKFFINPT
 Sbjct: 180 TVEYDEHDKPVRVDTVVISTQHDPNVTNDQIRQDVIEKVIKAVIPADYLDQTKFFINPT 239
 Query: 241 GRFVIGGPQGD SGLTGRKIIIVDTYGGYSRHGGGAFSGKD ATKVDRSASAYAARYIAKNIVA 300
 GRFVIGGPQGD SGLTGRKIIIVDTYGGYSRHGGGAFSGKD ATKVDRSASAYAARYIAKN+VA
 Sbjct: 240 GRFVIGGPQGD SGLTGRKIIIVDTYGGYSRHGGGAFSGKD ATKVDRSASAYAARYIAKNLVA 299
 Query: 301 ADLAKKVEVQLAYAIQVAPVSVRVDTFGTGVIAEADLEAAVRQIFDLRPAGIINMLDLK 360
 A L K EVQLAYAIQVAPVSVRVDTFGT + EA LEAAVRQ+FDLRPAGII MLDLK
 Sbjct: 300 AGLVTKEVQLAYAIQVAPVSVRVDTFGTSTVPEAVLEAAVRQVFDLRPAGIIMLDLK 359
 Query: 361 RPIYRQTAAAYGHMGRDIDLPWERVDKVQALKDFI 395
 RPIY+QTAAAYGHMGRDIDLPWER++KV AL + +
 Sbjct: 360 RPIYKQTAAAYGHMGRDIDLPWERLNKVDALEAV 394

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1593-

Example 1444

A DNA sequence (GBSx1530) was identified in *S.agalactiae* <SEQ ID 4429> which encodes the amino acid sequence <SEQ ID 4430>. This protein is predicted to be a transcriptional repressor of the biotin operon. Analysis of this protein sequence reveals the following:

```

5   Possible site: 24
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.16    Transmembrane  188 - 204 ( 188 - 204)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

15 A related GBS nucleic acid sequence <SEQ ID 9755> which encodes amino acid sequence <SEQ ID 9756> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB05404 GB:AP001512 transcriptional repressor of the biotin
  operon [Bacillus halodurans]
  Identities = 102/315 (32%), Positives = 169/315 (53%), Gaps = 18/315 (5%)

20  Query: 10  ILSKNNNFISGETMANQLNISRTAIWKGIKTLEELGLEIESVTNKGRLVSG-DILLPEQ 68
      +L+ ++F+SGE ++ + SRTA+WK I+ L + G E+E+V KGYR+V D + P
  Sbjct: 9  LLTAGDDDFVSGEKISQAIGCSRTAVWKHIEELRKSQYEVAVQKGYRIVKRPDQIKPHD 68

25  Query: 69  LE-----QEIGIKVSLNNSASTQLDAKMGIESKLKTPHLFLAPNQKKAKGRFDRPFPTS 123
      ++ + G +++ ++ASTQ A + K H+ LA Q KGR R +++
  Sbjct: 69  IQVLETERFPGREITYLESTASTQTVALKLAQEGAKEGHIVLANEQTSQKGRMGWYSP 128

30  Query: 124 NQGGIYMSLLQLPNVPIDIKPYTMVAVSAVKASRLTGITPEIKWVNDIYLDNKKIAG 183
      I MS++ +P +P + T++ A + V+AI TG+ +IKW ND+ +D KKI G
  Sbjct: 129 PGSSISMSIIFRPQLPPQKAPQITLLTAVAIVRAIKETTGLDSQIKWPNLLIDGKKIVG 188

35  Query: 184 ILTEAIASVESGLVTNVIIGLGINFYIKE--FPRALTKRAGSLFTEQ-PTITRNQLITEI 240
      ILTE A +S V +VI G+GIN +E F + K A SL ++ I R LI I
  Sbjct: 189 ILTEMQADQDS--VHSVVIQIGIGINVNHQEEAFABEIRKIATSLAIKKGEPIQRAPLIAAI 246

40  Query: 241 W---NLFFNIPLDHLK----VYREKSLVLDRTVSFMDGQTMYSKKAIDITDKGYLVVEL 293
      LF+++ L+ ++ ++ + + + + G A ITD G L++E
  Sbjct: 247 LKNIELFYDLYLQHGFSSRIKPLWEAHAISIGKRIRARMLNDVKFGVAKGITDDGVLLLED 306

  Query: 294 DDGQLKTLRSGEISL 308
      DDG+L ++ S +I +
  Sbjct: 307 DDGKLHSIYSADIEI 321

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4431> which encodes the amino acid sequence <SEQ ID 4432>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 34
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.49    Transmembrane  194 - 210 ( 194 - 211)

   ----- Final Results -----
      bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

55 The protein has homology with the following sequences in the databases:

```

>GP:BAB05404 GB:AP001512 transcriptional repressor of the biotin
  operon [Bacillus halodurans]
  Identities = 98/315 (31%), Positives = 165/315 (52%), Gaps = 18/315 (5%)

```

-1594-

Query: 10 LLSQTDDFVSGEYLADQLSISRTSVWKSISLENQGIQIDSLKHKGYRMVQG-DILLPKT 68
 LL+ DDFVSGE ++ + SRT+VWK I+ L G ++++++ KGYR+V+ D + P
 Sbjct: 9 LLTAGDDFVSGEKISQAIGCSRTAVWKHIEELRKSQYVEAVQRKGYRIVKRPDQIKPHD 68

Query: 69 I-----SQGLGMPVTYTPHSQSTQLDAKQIEAHNSAPRLYLAPSQEAAGRLDRQFFSA 123
 I ++ G +TY + STQ A + + + LA Q + KGR+ R ++S
 Sbjct: 69 IQVVLETERFGRITYLESTASTQTVALKLAQEGAKEGHIVLANEQTSKGKRMGRGWYSP 128

Query: 124 STGGIYMSMYLKPNVPYADMPPTYMMVASSIVKAISRLTGIDTEIKWVNDIYLGNHKVG 183
 I MS+ +P +P P T++ A +IV+AI TG+D++IKW ND+ + K+ G
 Sbjct: 129 PGSSISMSIIFRQLPPQKAPQLTLLTAVAIIVRAIKETTGLDSIKWPNDDLIDGKKIVG 188

Query: 184 ILTEAITSVETGLITDVIIGVGLNFFVTD--FPEAIAQKAGSLFTEK-PTITRNDLIIDI 240
 ILTE + + VI G+G+N + F E I + A SL +K I R LI I
 Sbjct: 189 ILTE--MQADQDSVHSVIQIGIGINVNHQEEAFABEIRKIATSLAIKKGEPIQRAPLIAAI 246

Query: 241 WK-----LFLSIPVKDHSVYKEKSLVLNKQVTFIENSQEKRAIAIDLTDQGHILIVQF 293
 K L+L +++ ++ + K++ + K +A +TD G L+++
 Sbjct: 247 LKNIELFYDLYLQHGFSRIKLWEAHASIGKRIRARMLNDVKGVAKGITDDGVLLLED 306

Query: 294 ENGLDQTLRSGEISL 308
 ++G L ++ S +I +
 Sbjct: 307 DDGKLHSIYSADIEI 321

An alignment of the GAS and GBS proteins is shown below.

Identities = 191/311 (61%), Positives = 257/311 (82%)

Query: 1 MKTYEKIVQILSKNNNFISGETMANQNLNISRTAIWKGIKTLEELGLEIESVTNKGYRLVS 60
 MKT EKIYQ+LS+ ++F+SGE +A+QL+ISRT++WK IK+LE G++I+S+ +KGYR+V
 Sbjct: 1 MKTSEKIYQLLSQTDDFVSGEYLADQLSISRTSVWKSISLENQGIQIDSLKHKGYRMVQ 60

Query: 61 GDILLPEQLEQEIGIKVSLNNNSASTQLDAKMGIESKLKTPHLFLAPNQKKAKGRFDRPF 120
 GDILLP+ + Q +G+ V+ +S STQLDAK GIE+ P L+LAP+Q+ AKGR DR F
 Sbjct: 61 GDILLPKTISQGLGMPVTYTPHSQSTQLDAKQIEAHNSAPRLYLAPSQEAAGRLDRQF 120

Query: 121 FTSNQGGIYMSLLLPNPVPIEDIKPYTMVASSAVKAISRLTGITPEIKWVNDIYLDNKK 180
 F+++ GGIYMS+ L+PNVP D+ PYT+MVASS VKAISRLTGI EIKWVNDIYL N K
 Sbjct: 121 FSASTGGIYMSMYLKPNVPYADMPPTYMMVASSIVKAISRLTGIDTEIKWVNDIYLGNHK 180

Query: 181 IAGILTEAIAVESGLVTNVIIGLGINFYIKEFPRALTKRAGSLFTEQPTITRNQLITEI 240
 +AGILTEAI SVE+GL+T+VIIG+G+NF++ +FP A+ ++AGSLFTE+PTITRN LI +I
 Sbjct: 181 VAGILTEAITSVETGLITDVIIGVGLNFFVTDFFPEAIAQKAGSLFTEKPTITRNDLIIDI 240

Query: 241 WNLFFNIPLEDHLKVYREKSLVLDRTVSFMDGQIMYSGKAIDITDKGYLVVELDDGQLKT 300
 W LF +IP++DH+KVY+EKSLVL++ V+F++ AID+TD+G+L+V+ ++G L+T
 Sbjct: 241 WKLLSIPVKDHSVYKEKSLVLNKQVTFIENSQEKRAIAIDLTDQGHILIVQFENGDLQT 300

Query: 301 LRSGEISLSSW 311
 LRSGEISLSSW
 Sbjct: 301 LRSGEISLSSW 311

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1445

A DNA sequence (GBSx1531) was identified in *Sagalactiae* <SEQ ID 4433> which encodes the amino acid sequence <SEQ ID 4434>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -2.76 Transmembrane 3 - 19 (3 - 20)
 ----- Final Results -----

-1595-

```

bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1446

10 A DNA sequence (GBSx1532) was identified in *S.agalactiae* <SEQ ID 4435> which encodes the amino acid sequence <SEQ ID 4436>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -2.28    Transmembrane    24 - 40 ( 24 - 40)
15
----- Final Results -----
    bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4437> which encodes the amino acid sequence <SEQ ID 4438>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -1.91    Transmembrane    58 - 74 ( 58 - 75)
25
----- Final Results -----
    bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
30    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 37/67 (55%), Positives = 54/67 (80%), Gaps = 3/67 (4%)
35
Query: 1   MTKRQFIFMALLCSFETYFFNQSVMDGSWIFAI FWGVLLRLDLQKVYAISKFTKELIK-- 58
          MT RQF+FMA +C+FETYFFN ++ G+++FA+FWG+LL RDL++V+ I++ TK ++K
Sbjct: 36  MTIRQFLFMAFVCAFETYFFNDLLLSGNYLFALFWGLLLFRDLRRVHTINQLTKTILKTA 95
40
Query: 59  -STKKKD 64
          S KKID
Sbjct: 96  NSPKKKD 102

```

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1447

A DNA sequence (GBSx1533) was identified in *S.agalactiae* <SEQ ID 4439> which encodes the amino acid sequence <SEQ ID 4440>. This protein is predicted to be DNA polymerase III, gamma subunit (dnaZX). Analysis of this protein sequence reveals the following:

```

Possible site: 60
>>> Seems to have no N-terminal signal sequence
50

```

-1596-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1567(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4441> which encodes the amino acid sequence <SEQ ID 4442>. Analysis of this protein sequence reveals the following:

Possible site: 60

10 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 232 - 248 (232 - 249)

----- Final Results -----

15 bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 408/558 (73%), Positives = 473/558 (84%), Gaps = 6/558 (1%)

20 Query: 1 MYQALYRKYRSQTFDEMVGQSVISITTLKQAVSSKKISHAYLFSGPRGTGKTSAAKIFAKA 60
 MYQALYRKYRSQTFDEMVGQSVISITTLKQAV S KISHAYLFSGPRGTGKTSAAKIFAKA
 Sbjct: 1 MYQALYRKYRSQTFDEMVGQSVISITTLKQAVESGKISHAYLFSGPRGTGKTSAAKIFAKA 60

25 Query: 61 MNCNPQINGEPCNHCIDICRDITNGSLEDVIEIDAASNNGVDEIRDIRDKSTYAPSRATYK 120
 MNCNPQ++GEP CN C D I C R D I T N G S L E D V I E I D A A S N N G V D E I R D I R D K S T Y A P S R A T Y K
 Sbjct: 61 MNCNPQVDGEP CN Q C D I C R D I T N G S L E D V I E I D A A S N N G V D E I R D I R D K S T Y A P S R A T Y K 120

30 Query: 121 VYIIDEVHMLSTGAFNALLKTLLEPTENVVFILATTELHKIPATILSRVQRFEFKAIKLL 180
 VYIIDEVHMLSTGAFNALLKTLLEPTENVVFILATTELHKIPATILSRVQRFEFKAIK
 Sbjct: 121 VYIIDEVHMLSTGAFNALLKTLLEPTENVVFILATTELHKIPATILSRVQRFEFKAIKQK 180

35 Query: 181 AIRDHLAQILDKEAISYDLDALTLVARRAEGGMRDALSILDQALS LAKDNHISLDVAEEI 240
 AIR+HLA +LDKE I+Y++DAL L+ARRAEGGMRDALSILDQALS L+ DN +++ +AEEI
 Sbjct: 181 AIREHLAWVLDKEGIAYEVDALNL IARRAEGGMRDALSILDQALS LSPDNQVAIAIAEEI 240

40 Query: 241 TGSISLSAIDDYVSNILAHDTTEALAKLEVIFDSGKSMRSR FATDLLMYLRDLLLVVQAGGE 300
 TGSIS+ A+ DYV + T+ALA LE I+DSGKSMRSR FATDLL YLRDLLLVV+AGG+
 Sbjct: 241 TGSISILALGDYVRYVSQEQATQALAALETIYDSGKSMRSR FATDLLTYLRDLLLVVQAGGD 300

45 Query: 301 DSHSSDTFIANLNVKQDILFEMIDKVTSLPEIKNGSHPKVYAEMMTIQLSEMVEKNSS- 359
 + S F NL++ D +F+MI VTS LPEIK G+HP++YAEMMTIQL++ + S
 Sbjct: 301 NQRQSAVFDTNLSLSIDRIFQMIVVTSHLPEIKKGTHPRIYAEMMTIQLAQKEQILSQV 360

50 Query: 360 NIPADVTAE LDSLRLRELKSLKNEMS QL- SRADQSSSTQKVKVNNKTF TFKVDR TKILT IM 418
 N+ ++ +E+++L+ EL LK ++SQL SR D + + K K KT +++VDR IL IM
 Sbjct: 361 NLSGELISEIETLKNELAQLKQQLSQLSQSRPDSLARS DTK- -PKTTSYRVDRVTILKIM 418

55 Query: 419 EETVVD SQR SREYLEALKSAWNEILDNITAQDRALLMGSEPVLANSENAILAFDAAFNAE 478
 EETV +SQ+SR+YL+ALK+AWNEILDNI+AQDRALLMGSEPVLANSENAILAF+AAFNAE
 Sbjct: 419 EETVRNSQSRQYLDALKNWNEILDNISAQDRALLMGSEPVLANSENAILAFEAFAFNAE 478

60 Query: 479 QAMKRTDLNDIFGNIMSKAAGFSPN I L A V P R N D F N Q I R S D F A K K M K A Q K - - T E T E P E V N H 536
 Q M R +LND+FGNIMSKAAGFSPN I L A V P R D F I R +FA++MK+QK + E EV
 Sbjct: 479 QVMSRNNLNDMFGNIMSKAAGFSPN I L A V P R T D F Q H I R K B F A Q M K S Q K D S V Q E E Q E V A L 538

Query: 537 QIPEDFSYLAERIAIVED 554
 IPE F +L ++I ++D
 Sbjct: 539 DIPEGFD FLLDKINTIDD 556

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1597-

Example 1448

A DNA sequence (GBSx1534) was identified in *S.agalactiae* <SEQ ID 4443> which encodes the amino acid sequence <SEQ ID 4444>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence (or aa 1-19)

----- Final Results -----

bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06927 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 67/143 (46%), Positives = 96/143 (66%)

Query: 8 ENYQLLLQLQAQALFSDETNALANLSNASAMLNAMLPNSVFTGFYLFIDGEELILGPFQGGV 67
 E Y L+ Q AL E++A+ANL+NASA+L L + GFYL EL+LGPFQG
 Sbjct: 13 EKYSLVTKQLAALLEGESDAIANLANASALLYHFLEEVNWWGVFYLIKEGELVLGPFQGLP 72

Query: 68 SCVHITLGKGVCGESAQTAKTLIVDDVTKHANYISCDKAMSEIVVPMFKNGKLLGVLDL 127
 +CV I +G+GVC +A+ +T+ V+DV + +I+CD+ + SEIV+P+F+NG L GVLD+
 Sbjct: 73 ACVRPIGRGVCGTAKEEQTVRVEDVHQFPGHIACDAASRSEIVIPLFQNGVLYGVLDI 132

Query: 128 DSSLVADYDEIDQEYLEKFVGIL 150
 DS + + E +Q LE FV +L
 Sbjct: 133 DSPSLNRFSEEEQALLESFVDVL 155

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4445> which encodes the amino acid sequence <SEQ ID 4446>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.1753 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 122/164 (74%), Positives = 144/164 (87%)

Query: 1 MNKSKKIENYQLLLQLQAQALFSDETNALANLSNASAMLNAMLPNSVFTGFYLFIDGEELIL 60
 MNKSKKIE YQL++ QA+ LF++E+NALANLSNASA+LN LPNSVFTGFYLFDG+ELIL
 Sbjct: 1 MNKSKKIEYQLMIAQAKELFANESNALANLSNASALLNMTLPNSVFTGFYLFDGQELIL 60

Query: 61 GPFQGGVSCVHITLGKGVCGESAQTAKTLIVDDVTKHANYISCDKAMSEIVVPMFKNGK 120
 GPFQG VSCVHI LGKGVCGESAQ+ +T+I++DV +HANYISCD+ AMSEIVVPM K G
 Sbjct: 61 GPFQGRVSCVHIKLGKGVCGESAQSRRTIIINDVKQHANYISCDAAAMSEIVVPMVKEGH 120

Query: 121 LIGVLDLSDSSLVADYDEIDQEYLEKFVGILVEHTIWNLDMGVE 164
 L+GVLDLSDSSLVADYDE+DQEYLE FV + +E T + +MFGV+
 Sbjct: 121 LIGVLDLSDSSLVADYDEVDQEYLEAFVDLFLEKTFTFTFMFGVK 164

SEQ ID 4444 (GBS282) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 9; MW 19.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 6; MW 44.8kDa) and in Figure 63 (lane 7; MW 47kDa).

-1598-

The GBS282-GST fusion product was purified (Figure 211, lane 4; see also Figure 225, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 269), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1449

A DNA sequence (GBSx1535) was identified in *S.galactiae* <SEQ ID 4447> which encodes the amino acid sequence <SEQ ID 4448>. This protein is predicted to be uridine kinase (udk). Analysis of this protein sequence reveals the following:

```

10   Possible site: 24
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
15   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20   >GP:CAB14675 GB:Z99117 uridine kinase [Bacillus subtilis]
    Identities = 133/207 (64%), Positives = 167/207 (80%)

    Query: 1  MRKKPIIIGVTGGSGGKTSVSRILSNFPDQKITMIEHDSYYKQSHLTFEERVKTNYD 60
           M K P++IG+ GGSG GKTSV+R+I  F  I MI+ D YYKQSHL FEER+ TNYD
    Sbjct: 1  MGKNPVVIGIAGGSGGKTSVTRSIYEQFKGHSILMIQQDLYYKQSHLPFEERLNTNYD 60

25   Query: 61  HPLAFDTNLMIEQLNELIEGRPVDIPVYDYTKHTRSDRTIROEPQDVIIVEGILVLEDQR 120
           HPLAFD + +IE + +L+  RP++ P+YDY HTRS+ T+  EP+DVII+EGILVLED+R
    Sbjct: 61  HPLAFDNDYLIEHIQDLLNYRPIEKPIYDYKLHTRSEETVHVEPKDVIILEGILVLEDKR 120

30   Query: 121 LRDLMIDIKLFVDTDDDIRIIRIKRDMEEEDRSLSIIEQYTEVVKPMYHQFIEPTKRYA 180
           LRDLMIDIKL+VDTD D+RIIRRI RD+ ER RS+DS+IEQY  VV+PM++QF+EPTKRYA
    Sbjct: 121 LRDLMIDIKLYVDTDADLRIIRIMRDINERGRSIDSIEQYVSVVRPMHNFVEPTKRYA 180

35   Query: 181 DIVIPEGVSNIVAIIDLINTKVASILNE 207
           DI+IPEG N VAIDL+ TK+ +IL +
    Sbjct: 181 DIIIEGGQNHVAIDLMVTKIQTILEQ 207

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4449> which encodes the amino acid sequence <SEQ ID 4450>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 39
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9151> which encodes the amino acid sequence <SEQ ID 9152>. Analysis of this protein sequence reveals the following:

```

50   Possible site: 35
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
55   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```


-1599-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 173/207 (83%), Positives = 193/207 (92%)

5 Query: 1 MRKKPIIIGVTGGSGGKTSVSRILSNFPDQKITMIEHDSYYKDQSHLTFEERVKTNYD 60
 M KKPIIIGVTGGSGGKTSVSRIL +FP+ +I MI+HDSYYKDQSH++FEERVKTNYD
 Sbjct: 5 MLKKPIIIGVTGGSGGKTSVSRILDSFPNARIAMIQHDSYYKDQSHMSFEERVKTNYD 64

10 Query: 61 HPLAFDTNLMIEQLNELIEGRPVDPVYDYTKHTRSRTIRQEPQDVIIIVEGILVLEDQR 120
 HPLAFDT+ MI+QL EL+ GRPVDIP+YDY KHTRS+ T RQ+PQDVIIIVEGILVLED+R
 Sbjct: 65 HPLAFDTDFMIQQLKELLAGRVPDIPIYDYKKHTRSNTTFRQDPQDVIIIVEGILVLEDER 124

15 Query: 121 LRDLMDIKLFVDTDDDIRIIRRIKRDMERDRSLDSIEQYTEVVKPMYHQFIEPTKRYA 180
 LRDLMDIKLFVDTDDDIRIIRRIKRDM ER RSL+SII+QYT VVKPMYHQFIEP+KRYA
 Sbjct: 125 LRDLMDIKLFVDTDDDIRIIRRIKRDMMERGRSLESIIDQYTSVVKPMYHQFIEPSKRYA 184

Query: 181 DIVIEGVSNIWAIDLINTKVASILNE 207
 DIVIEGVSNI+VAID+IN+K+ASIL E
 20 Sbjct: 185 DIVIEGVSNNVAIDVINSKIASILGE 211

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1450

25 A DNA sequence (GBSx1536) was identified in *Sagalactiae* <SEQ ID 4451> which encodes the amino acid sequence <SEQ ID 4452>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5083(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12572 GB:Z99108 similar to RNA helicase [Bacillus subtilis]
 Identities = 140/343 (40%), Positives = 202/343 (58%), Gaps = 9/343 (2%)

40 Query: 10 QDKLTQRQFDDLDIIONKLFQPIITDGDNIIGISPTGTGKTLAYLFPTLLKLQPK-KSQQL 68
 Q+ F T +Q + Q I DG +++ SPTGTGKTLAY P L +++P+ K Q
 Sbjct: 16 QENWNASGFQKPTPVQEQAAQLIMDKDVIAESPTGTGKTLAYALFVLERIKPEQKHPQA 75

45 Query: 69 LILAPNSELAQIFDVTKEWAEPLGLTAQLFLSGSSQKRQIERLKKGPEILIGTAGRVFE 128
 +ILAP+ EL QIF V ++W L A + G++ K+Q+E+LKK P I++GT GRVFE
 Sbjct: 76 VILAPSRRLVMQIFQVIQDWKAGSELRAASLIGGANVKKQVEKLKKPHIIVGTPGRVFE 135

50 Query: 129 LVKLKKIKMMNINTIVLDEFDELLGDSQYHFVDNIINRVPRDQQMIYISATNKLDNS--- 185
 L+K KK+KM + TIVLDE D+L+ + II RD+Q++ SAT K +
 Sbjct: 136 LIKAKKLMHEVKTIVLDETDQLVLPHERETMTKQIIKTTLRDRQLLCFSATLKKETEDVL 195

55 Query: 186 -KLADNTITIDLSNQKLDL--IKHYITVDKRERTDLLRKFSNIPDFRGLVFFNLSLDLG 242
 +LA + + K + +KH Y+ D+R++ LL+K S + + LVF + +L
 Sbjct: 196 RELAQEPEVLKVQRSKAEAGVKHQYLIQDQDKVLLQKLSRLEGMQALVFVRDIGNLS 255

60 Query: 243 ACEERLQFNRAVASLASDINIKFRKVILEKFRNHDISLLGLTDLVARGIDIDNLEYVIN 302
 E+L ++ L S+ R I+ F++ + LLL TD+ ARG+DI+NL YVI+
 Sbjct: 256 VYAEKLAYHHVELGVHSEAKMERAKIIATFEDGEFPLLLATDIAARGLDIENLPYVIH 315

Query: 303 FDIARDKETTYTHRSGRTRMGKEGCVITFVTHKEELKQLKKYA 345
 DI D++ Y HRSGRTRG KKEG V++ VT EE K LKK A
 Sbjct: 316 ADIP-DEGYYVHRSGRTRGAGKEGNVLSLVTKLEESK-LKKMA 356

-1600-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4453> which encodes the amino acid sequence <SEQ ID 4454>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3847(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 273/358 (76%), Positives = 312/358 (86%)

```

15 Query: 1  MITKFPDQWQDKLTQRQFDDLTDIQNKLFQPIITDGDNILGISPTGTGKTLAYLFPTLLKL 60
      MITKFP QWQ+KL Q F LT IQ + FQPI DG N LGISPTGTGKTLAY+FP LL L
Sbjct: 12  MITKFPQWQEKLDQVAFTHLTPIQEQAFQPIVDGKNFLGISPTGTGKTLAYVFPNLLAL 71

20 Query: 61  QPKKSQQLLILAPNSELAGQIFDVTKEWAEPLGLTAQLFLSGSSQKRQIERLKKGP EILI 120
      PPKKSQQLLILAPN+ELAGQIF+VTK+WA+PLGLTAQLF+SG+SQKRQIERLKKGP EILI
Sbjct: 72  TPKKSQQLLILAPNTELAGQIFEVTKDWAQPLGLTAQLFISGTSQKRQIERLKKGP EILI 131

25 Query: 121 GTAGRVFELVKKKKIKMMNINTIVLDEFDELLGDSQYHFVDNIINRVPRDQMIYISATN 180
      GT GR+FEL+KLKKIKMM++NTIVLDE+DELLGDSQY FV I + VPRD QM+Y+SATN
Sbjct: 132 GTPGRIFELIKLKKIKMMSVNTIVLDEYDELLGDSQYDFVQKISHYVPRDHQMVYMSATN 191

30 Query: 181 KLDNSKLADNTITIDLSNQKLDTIKHYYITVDKRERTDLLRKFSNIPDFRGLVFFNLSLSD 240
      K+D + LA NT IDLS Q D I+H+Y+ VDKRERTDLLRK F+NIP FR LVFFNLSLSD
Sbjct: 192 KVDQTS LAPNTFCIDLSEQTNDAIQH FYLMVDKRERTDLLRKFTNIPHFRLVFFNLSLSD 251

35 Query: 241 LGACEERLQFNRAVSLASDINIKFRKVILEKFKNHDISLLLGTDLVARGIDIDNLEYV 300
      LGA EERLQ+N A+AVSLASDIN+KFRK ILEKFK+H +SLLL TDLVARGIDIDNL+YV
Sbjct: 252 LGATEERLQYNGAAVSLASDINVKFRKTILEKFKSHQLSLLLATDLVARGIDIDNLDYV 311

40 Query: 301 INFDIARDKETYTHRSRGRTGRMGKEGCVITFVTHKEELKQLKKYATVTELVLHNQKLH 358
      I+FD+ARDKE YTHR+GRTGRMGK G VITFV+H E+LK+LKK+A V+E+ L NQ+LH
Sbjct: 312 IHFDVARDKENYTHRAGRTGRMGKSGIVITFVSHPEDLKKLKKFAKVSEISLKNQQLH 369

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1451

A DNA sequence (GBSx1537) was identified in *S.agalactiae* <SEQ ID 4455> which encodes the amino acid sequence <SEQ ID 4456>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -1.38 Transmembrane 15 - 31 (13 - 31)

----- Final Results -----

bacterial membrane --- Certainty=0.1553(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1601-

Example 1452

A DNA sequence (GBSx1538) was identified in *S.agalactiae* <SEQ ID 4457> which encodes the amino acid sequence <SEQ ID 4458>. This protein is predicted to be peptidoglycan GlcNAc deacetylase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 28
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -8.92    Transmembrane    4 - 20 ( 1 - 26)

----- Final Results -----
10      bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAB96552 GB:AJ251472 peptidoglycan GlcNAc deacetylase
      [Streptococcus pneumoniae]
      Identities = 133/431 (30%), Positives = 228/431 (52%), Gaps = 20/431 (4%)

20   Query: 5   IIGIFSLIIIIAILAWQGFSLKHK--BIKLQAVVEKEIRIAEKTVEVVKRQKTERVLFL 62
      +IGI ++ I + + F + K E K++ EK+ +++E + RQ V+
      Sbjct: 21 LIGILAIISICLLGGFIAFKIYQQKSFQKIESLKEKDDQLSEGNQKEHFRQGGQAEVIAY 80

      Query: 63 EPKGYDKSLSADILKWNQKSFEHKKFYDNQYIIILRPQLADSNFANVKLSIYQILYQKEK 122
      P +K +S+ NQ + + DN Q +S V ++ ++Y
25   Sbjct: 81 YPLQGEKVISSVRELINQDVKDKLESKDNLVFYYTEQ--EESGLKGVNVRNVTKQIYDLVA 139

      Query: 123 GSMFQKSSRLLRTYLLDQNKPPFELDELLAHNISGFKAILLENIAPGTQLK--EHDSNKEF 180
      + + L L ++ +PF LD+L + + +++ + + K E D +++
30   Sbjct: 140 FKIEETKTSLSGKVLHLEDGQPFTLDQLFSDASKAKEQLIKELTSFIEDKKIEQDQSEI 199

      Query: 181 LKTGRVTD----GLDVKDGKLI-----NDLKLPLDKLYNVIDESYLKSSDLDLVS 227
      +K D D KD ++I+ ++ LP+ ++VI SYL D L
      Sbjct: 200 VKNFSDQDLSAWNFDYKDSQIILYPSPVVENLEEIALPVSAFFDVIQSSYLEKDAALYQ 259

35   Query: 228 NLKAKAPR--VALTFDDGPNKTTTPKALEILKRYNAKATFFVMGQSAVGHTDILQRMHAE 285
      + K + VALTFDDGPN TTP+ LE L +Y+ KATFFV+G++ G+ D+++R+ +E
      Sbjct: 260 SYFDKKHQKVVALTFDDGPNPATTTPQVLETLAKYDIKATFFVLGKNVSGNEDLVKRIKSE 319

      Query: 286 GHEIGNHTWDHPNLTKLPAEKIKEEIHKTNDLIMKATGQKPVYLRPPYGATNATVKTVTG 345
      GH +GNH+W HP L++L ++ K++I T D++ K G +RPPYGA ++
40   Sbjct: 320 GHVVGNHWSWHPILSQLSLDEAKKQITDTEVLTKVLGSSSKLMRPPYGAITDDIRNSLD 379

      Query: 346 LKEMLWSVDTEDWKNHNTQAMMTNIKKQLRPGGVILMHDIHQTTIDALPTIMDYLTQGY 405
      L ++W VD+ DWK+ N +++T I+ Q+ G ++LMHDIH T++ALP +++YL QGY
45   Sbjct: 380 LSFIMWDVDSLWKSNEASILTEIQHQVANGSIVLMHDIHSPTVNALPRVIEYLNQGY 439

      Query: 406 YFVTVGELYST 416
      FVT+ E+ +T
50   Sbjct: 440 TFFVTIPEMLNT 450

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4459> which encodes the amino acid sequence <SEQ ID 4460>. Analysis of this protein sequence reveals the following:

```

55   Possible site: 22
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -12.58    Transmembrane    6 - 22 ( 1 - 27)

----- Final Results -----
60      bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-1602-

The protein has homology with the following sequences in the databases:

1GB:AJ251472 peptidoglycan GlcNAc deacetylase [Strep... 239 4e-62

>GP:CAB96552 GB:AJ251472 peptidoglycan GlcNAc deacetylase
[Streptococcus pneumoniae]

Identities = 136/438 (31%), Positives = 230/438 (52%), Gaps = 23/438 (5%)

Query: 3 KLNVLVGLLSILMLSLAI----VFINRWKLNEDSQRIVLAEKKKNTSDLVKAVKHIKK 58
K +L+ L+ IL +S+ + + ++ Q+I +K+K+ +H ++

10 Sbjct: 13 KTRHVLALIGILAISICLLGGFIAFKIYQQKSFEQKIESLKKKDDQLSEGNQKEHFRQ 72

Query: 59 DQKDYFFSPK--QADDFVDNLPVSLYKKKNSDKELILVRPKLQSSHLRSVNTLTISK 116
Q + + P++ + + + + K S L+ + + S L+ V ++K

15 Sbjct: 73 GQAEVIAYPLQGEKVISSVRELINQDVKDKLESKDNLFVFFYTEQESGLKGVVNRNVTK 132

Query: 117 IVYQKKFFHLAKSEKIVSTYHVTDDLKPFQVKDLVSGHL---ERIQEEVEKKYPDAGFN 173
+Y F + + + + H+T+D +PF + L S E++ +E+ D

20 Sbjct: 133 QIYDLVAFKIEETKTSLGKVLHTEGQPFITLDQLFSDASKAKELIKELTSFIEDKKIE 192

Query: 174 SDKYNGLKESNS---LLSDGFEVKSGNLIFD-----KKLTIPLTTLFDVINPDPLAN 222
D+ + ++ S L + F+ K +I +++ +P++ FDVI +L

25 Sbjct: 193 QDQSEQIVKNFSDQDLAWNFYKDSQIILYPSFVVENLEEIALPVSAFFDVIQSSYLLE 252

Query: 223 SDRAAYDNYRITYKEQHPKKLVALTFDDGPDPTTTPQVLDILAKYQAKGTFMIGSKVNN 282
D A Y +Y K Q K+VALTFDDGP+P TTPQVL+ LAKY K TFF++G V N

30 Sbjct: 253 KDAALYQSYFDKKHQ---KVVALTFDDGPNPATTPQVLETLAKYDIKATFFVLGKNVSGN 309

Query: 283 ENLTKRVSDAGHEIANHTWDHFNLTNLSVSEIQHVNMTNQAIEKACGKKPRYLPPYGA 342
E+L KR+ GH + NH+W HP L+ LS+ E + Q+ T + K G + +RPPYGA

35 Sbjct: 310 EBLVKRIKSEGHVVGNSHWSHPILSLSLDEAKKQITDTEVDLTKVLGSSSKLMRPPYGA 369

Query: 343 TNATVQSSGLTQMLWTVTRDWNHSTDGIMTNVKNQLQPGGVVLMHDIHQTTINALPT 402
++ S L+ ++W VD+ DW++ + I+T +++Q+ G +VLMHDIH T+NALP

40 Sbjct: 370 ITDDIRNSLDLSFIMWDVDSLWKSNEASILTEIQHQQVANGSIVLMHDIHSPTVNALPR 429

Query: 403 VMEYLKAEGYECVTVSEL 420
V+EYLK +GY VT+ E+

45 Sbjct: 430 VIEYLKNQGYTFVTIPEM 447

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 169/420 (40%), Positives = 259/420 (61%), Gaps = 12/420 (2%)

Query: 4 LIIGIFSLIIAILAWQGFSLKHKEIKLQQAQVVEKEIRIAEKTVEVVKRQKTER--VLF 61
+++G+ S++++ LA + K E + + EK+ ++ ++ VK K ++ +

45 Sbjct: 7 ILVGLLSILMLS-LAIVFINRWKLNEDSQRIVLAEKKKNTSDLVKAVKHIKKDQKDYFF 65

Query: 62 LEPKGYDKLSADILKWNQKSFEHKKFYDNQYIILRPQLADSNFANVKLSIYQILYQKE 121
P D L S KK D + I++RP+L S+ +V L+I +I+YQK+

50 Sbjct: 66 FSPKQADDFVDNLP--VSLYKKKNSDKELILVRPKLQSSHLRSVNTLTISKIVYQKK 122

Query: 122 KGSFMQKSSRLRLTYLLDQNKPFELDELLAHNISGFKAILENIAPGTQLKEHDSNKEFL 181
+ +KS +++ TY + + KPF++ +L++ ++ + +E P N

55 Sbjct: 123 FFHLAKKSEKIVSTYHVTDDLKPFQVKDLVSGHLERIQQEEVEKKYPDAGFNSDKYNLKE 182

Query: 182 KTGRVTDGLDVKGDKLIIND-LKLPLDKLYNVIDESYLKSSDLDLVSNL---KAKAPR-- 235
++DG +VK G LI + L +PL L++VI+ +L +SD N K + P+

60 Sbjct: 183 SNSLLSDGFEVKSGNLIFDKKLTITPLTTLFDVINPDFLANSRAAYDNYRITYKEQHPKKL 242

Query: 236 VALTFDDGPNKETTTPKALEILKRYNAKATFFVMGQSAVGHTDILQRMHAEHGHEIGNHTWD 295
VALTFDDGP+ TTP+ L+IL +Y AK TFF++G V + ++ +R+ GHEI NHTWD

65 Sbjct: 243 VALTFDDGPDPTTTPQVLDILAKYQAKGTFMIGSKVNNENLTKRVSDAGHEIANHTWD 302

Query: 296 HPNLTKLPAEKIKEEIIHKTNLIMKATGQKPVYLRPPYGATNATVKTVTGLKEMLSVSDT 355
HPNLT L +I+ +++ TN I KA G+KP YLRPPYGATNATV+ +GL +MLW+VDT

65 Sbjct: 303 HPNLTNLSVSEIQHVNMTNQAIEKACGKKPRYLPPYGATNATVQSSGLTQMLWTVDT 362

-1603-

Query: 356 EDWKNHNTQAMMTNKKQLRPGGVILMHDHQTITDALPTIMDYLTTOGGYFVTVGELYS 415
 DW+NH+T +MTN+K QL+PGGV+LMHDHQTIT+ALPT+M+YL +GY VTV ELY+
 Sbjct: 363 RDWENHSTDGIMTNVKNQLQPGGVVLMHDHQTITINALPTVMEYLKAEGYECVTVSELYA 422

5 GBS281d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 8-10; MW 71.5kDa) and in Figure 187 (lane 10; MW 71kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 12; MW 46.5kDa) and in Figure 183 (lane 2; MW 46kDa). Purified GBS281d-GST is shown in lane 6 of Figure 237.

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1453

A DNA sequence (GBSx1539) was identified in *S.galactiae* <SEQ ID 4461> which encodes the amino acid sequence <SEQ ID 4462>. Analysis of this protein sequence reveals the following:

15 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.2488(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4463> which encodes the amino acid sequence <SEQ ID 4464>. Analysis of this protein sequence reveals the following:

25 Possible site: 40
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.2799(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 311/475 (65%), Positives = 389/475 (81%)

35 Query: 1 MTKEYQNYVNGEWKSSVNQIEILSPIDSSLGFPVAMPTRREEVDHAMKAGREALPAWAALT 60
 + K+Y+N VNGEWK S N+I I +P LG VPAMT+ EVD + ++AL W AL+
 Sbjct: 1 LAKQYKNLVNGEWKLSENEITTIYAPATGEELGSPVAMPQAEVDVAYASAKKALSDWRALS 60

40 Query: 61 VYERAQYLHKAADIIERDKKEIATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLST 120
 ERA YLHKAADI+ RD E+I +L+KE++K + A+V+EV+RTA++I YAAEEG+R+
 Sbjct: 61 YVERAAYLHKAADILVRDAEKIGAILSKEVAKGHKAAVSEVIRTAEIINYAAEEGLRMEG 120

45 Query: 121 SADEGGKMDASTGHKLAVIRRPQVGIIVLAIAPYNYPVNLGSKIAPALIGGNVVMFKPPT 180
 EGG +A++ K+A++RR+PVG+VLAI+P+NYPVNL+GSKIAPALI GNVV KPPT
 Sbjct: 121 EVLEGGSFEEASKKKIAIVRREPVGVLVAISPFPNYPVNLGSKIAPALITAGNVVALKPPT 180

50 Query: 181 QGSVSGLVLAFAFAEAGLPAGVFNTITGRGSEIGDYIVEHEEVNFINTGSTPVGKRIGK 240
 QGS+SGL+LA+FAFAEAG+PAGVFNTITGRGS IGDYIVEHE V+FINFTGSTP+G+ IGK
 Sbjct: 181 QGSI SGLLAEFAFAEAGIPAGVFNTITGRGSEIGDYIVEHEAVSFINTGSTPIGEIGIGK 240

55 Query: 241 LAGMRPIMLELGGKDAGVVLADADLDNAAKQIVAGAYDYSQQRCTAIKRVLVVEEVADEL 300
 LAGMRPIMLELGGKD+ +VL DADL AAK IVAGA+ YSGQRCTA+KRVLV+++VAD+L
 Sbjct: 241 LAGMRPIMLELGGKDAIVLEADLALAAKNIVAGAFGYSGQRCTAVKRVLVMDKVADQL 300

Query: 301 AEKISENVAKLSVGDPFDNATVTPVIDDNSADFIESLVVDARQKGAKELNEFKRDGRLLT 360

A +I V KLSVG P D+A +TP+ID ++ADF+E L+ DA KGA L F R+G L++
 Sbjct: 301 AAEIKTLVEKLSVGMPEDDADITPLIDTSAADFVEGLIKDATDKGATALTAFNREGNLIS 360
 Query: 361 PGLFDHVTLDMLKLAWEFPFGPILPIIRVKDAEEAVAIANKSDFGLQSSVFTRDFQKAFDI 420
 P LFDHVT DM+LAWEEPFGP+LPPIIRV EEA+ I+N+S++GLQ+S+P +F KAF I
 Sbjct: 361 PVLFDHVTDMRLAWEEPFGPVLPIIRVTTVEEAIKISNESEYGLQASIFTTNPFKAFGI 420
 Query: 421 ANKLEVGTVHINNKTGRGPDNFPFLGLKGSGAGVQGIIRYSIEAMTNVKSIVFDMK 475
 A +LEVGTVH+NNKT RG DNFPFLG K SGAGVQG++YSIEAMT VKS+VFD++
 Sbjct: 421 AEOLEVGTVHLNNKTORGDNFPFLGAKKSGAGVQGVKYSIEAMTIVKSIVFDDIQ 475

```

15 Lipop: Possible site: -1    Crend: 3
    McG: Discrim Score:      -15.11
    GvH: Signal Score (-7.5): 0.17
        Possible site: 57
    >>> Seems to have no N-terminal signal sequence
    ALOM program   count: 0 value:   1.22 threshold:  0.0
20     PERIPHERAL Likelihood = 1.22      187
        modified ALOM score: -0.74

    *** Reasoning Step: 3

25 ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.2488(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

Streptococcus mutans

EGAD|42413| NADP-dependent glyceraldehyde-3-phosphate dehydrogenase Insert characterized

EGAD|42413|110509 NADP-dependent glyceraldehyde-3-phosphate dehydrogenase Insert characterized

SP|Q59931|GAPN_STRMU NADP-DEPENDENT GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.9) (NON-PHOSPHORYLATING GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE) (GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE [NADP+]) (TRIOSEPHOSPHATE DEHYDROGENASE). Edit characterized

GP|642667|gb|AAA91091.1||L38521 NADP-dependent glyceraldehyde-3-phosphate dehydro Insert characterized

ORF01688(301 - 1725 of 2025)

EGAD|42413|44796(1 - 475 of 475) NADP-dependent glyceraldehyde-3-phosphate dehydrogenase {Streptococcus mutans}|EGAD|42413|110509 NADP-dependent glyceraldehyde-3-phosphate dehydrogenase {Streptococcus mutans}|SP|Q59931|GAPN_STRMU NADP-DEPENDENT GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.9) (NON-PHOSPHORYLATING GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE) (GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE [NADP+]) (TRIOSEPHOSPHATE DEHYDROGENASE).GP|642667|gb|AAA91091.1||L38521 NADP-dependent glyceraldehyde-3-phosphate dehydro

%Match = 49.3

%Identity = 66.7 %Similarity = 82.5

Matches = 317 Mismatches = 83 Conservative Sub.s = 75

55 195 225 255 285 315 345 375 405
*GLKNLYFFIESLDIVKFLRKICQIIEINR*SDRINLLQCKRRFTLTKEVQNYVNGEWKSSVNQIEILSPIDSSSLGFVP
:|:|:||||||| | |:| | : || |
MTKQYKNYVNGEWKLSENEIKIYEPASGAELGSVP
10 20 30

60 435 465 495 525 555 585 615 645
AMTREEVDHAMKAGREALPAWAALT'YVERAQYLHKAADIIEFDKEBIATVLAKEISKAYNASVTEVVRTADLIRYAAEEG
|:| |:|: : :| ||| |:| ||| |:| |:|:| : |:|:| | : |:|:|||||:| | |||||
AMSTEEVDYVYASAKKAQPAWRALS'YIERAAYLHKVADILMRDKEKIGAILSKEVAGKYKSAVSEVVRTAETINYAAEEG
50 60 70 80 90 100 110

-1606-

5
Query: 121 DNPYMQERAADIRDVAKRVLHLLGVKLPNPATINEESIVIAHDLTPSDTAQLNKQFVKA 180
Sbjct: 121 DNPYMQERAADIRDVAKRVLHLLGKLPNPATIDEESIVIAHDLTPSDTAQLNKQFVKA 180

10
Query: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITERVQDQGLIANGITGEVIIPT+ Q++ 240
Sbjct: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITSRVKGDIVANGITGEVIINPTDEQVAE 240

15
Query: 241 FKAAGEAYAKQKAEWALLKDAQTVTADGKHFELAAANIGTPKDVEGVNENGAEAVGLYRTE 300
Sbjct: 241 FKAAGEAYAKQKAEWALLKDAKTVTADGKHFELAAANIGTPKDVEGVNANGAEAVGLYRTE 300

20
Query: 301 FLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMNPFLGFR 360
Sbjct: 301 FLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPYLDLPKEMNPFLGFR 360

25
Query: 361 ALRISISSETGDMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFEEKANLLAD 420
Sbjct: 361 ALRISISSETGNAMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFDEEKANLKA 420

30
Query: 421 GVAVAEIEVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRMNEQVSILYQP 480
Sbjct: 421 GVAVSDDIQVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRMNEQVSILYQP 480

35
Query: 481 YNPSILRLINNVIKAAHAEGKWGMCGEMAGDQAVPLLVGMGLDEFSMSATS+LRTSL 540
Sbjct: 481 YNPSILRLINNVIKAAHAEGKWGMCGEMAGDQKAVPLLVGMGLDEFSMSATSILRTSL 540

40
Query: 541 MKKLD TAKMEEYANRALSECSTMEEVLELQKEYVDFD 577
Sbjct: 541 MKKLD TAKMEEYANRALTECSTMEEVLELSKEYVND 577

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4467> which encodes the amino acid sequence <SEQ ID 4468>. Analysis of this protein sequence reveals the following:

Possible site: 39
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0875(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 540/577 (93%), Positives = 561/577 (96%)

Query: 1 MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVEDTNAEEARLDVALQASQDELSVIRE 60
Sbjct: 1 MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTV DTNAEEARLDVALQA+QDELSVIRE 60

50
Query: 61 KAVESLGEEAAAVFDAHLMVLSDPEMINQIKETIRAKQVNAETGLKEVTDMFITIFEGME 120
Sbjct: 61 NAVESLGEEAAAVFDAHLMVLADPEMISQVKETIRAKQVNAETGLKEVTDMFITIFEGME 120

55
Query: 121 DNPYMQERAADIRDVAKRVLHLLGVKLPNPATINEESIVIAHDLTPSDTAQLNKQFVKA 180
Sbjct: 121 DNPYMQERAADIRDVAKRVLHLLGVKLPNPATINEESIVIAHDLTPSDTAQLNKQFVKA 180

60
Query: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITERVQDQGLIANGITGEVIIPT+ Q+ A 240
Sbjct: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITKRVKGDIVANGITGEVIIDPSEDQVLA 240

65
Query: 241 FKAAGEAYAKQKAEWALLKDAQTVTADGKHFELAAANIGTPKDVEGVNENGAEAVGLYRTE 300
Sbjct: 241 FKAAGEAYAKQKAEWALLKDAQTVTADGKHFELAAANIGTPKDVEGVNENGAEAVGLYRTE 300

-1607-

Query: 301 FLYMDSQDFPTEDEQEYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMNPFLGFR 360
 Sbjct: 301 FLYMDSQDFPTEDEQEYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMNPFLGFR 360

Query: 361 ALRISISSETGDAMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFEEKANLLAD 420
 Sbjct: 361 ALRISISSETGDAMFRTQ+RALLRASVHGQLRIMFPMVALLKEFRAAKA+F+EEKANLLA+ 420

Query: 421 GVAVAEGIEVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRMNEQVSYLEYQP 480
 Sbjct: 421 GVAVADDIQVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRMNEQVSYLEYQP 480

Query: 481 YNPSILRLINNVIAAHAEGKWAGMCGEMAGDQTAVPLLVGMGLDEFSMSATSVLRTRSL 540
 Sbjct: 481 YNPSILRLINNVIAAHAEGKWAGMCGEMAGDQ AVPLLVGMGLDEFSMSATSVLRTRSL 540

Query: 541 MKKLD+AKMEEYANRAL+ECST EEV+EL KEYV D 577
 Sbjct: 541 MKKLDSAKMEEYANRALTECSTAEVLELSKEYVSED 577

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1455

A DNA sequence (GBSx1541) was identified in *S.agalactiae* <SEQ ID 4469> which encodes the amino acid sequence <SEQ ID 4470>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to a protein from *S.bovis*:

>GP:BAA78048 GB:AB027569 histidine containing protein [Streptococcus bovis]
 Identities = 86/87 (98%), Positives = 87/87 (99%)

Query: 1 MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSLMGVMSLGVGQGAD 60
 Sbjct: 1 MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSLMGVMSLGVGQGAD 60

Query: 61 VTISAEGADADDAIAAIEETMTKEGLA 87
 Sbjct: 61 VTISAEGADADDAIAAIEETMTKEGLA 87

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4471> which encodes the amino acid sequence <SEQ ID 4472>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/87 (98%), Positives = 87/87 (99%)

-1608-

Query: 1 MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSINGVMSLGVGQGAD 60
 MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSINGVMSLGVGQGAD
 Sbjct: 1 MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSINGVMSLGVGQGAD 60

5 Query: 61 VTISAEGADADDAIAAIEETMTKEGLA 87
 VTISAEGADA+DAIAAIEETMTKEGLA
 Sbjct: 61 VTISAEGADAEDAIAAIEETMTKEGLA 87

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 1456

A DNA sequence (GBSx1542) was identified in *S.agalactiae* <SEQ ID 4473> which encodes the amino acid sequence <SEQ ID 4474>. This protein is predicted to be glutaredoxin-like protein nrhd (b2673). Analysis of this protein sequence reveals the following:

15 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.4532(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:CAA63372 GB:X92690 glutaredoxin-like protein [Lactococcus
 lactis]
 Identities = 42/70 (60%), Positives = 53/70 (75%)

Query: 4 ITVFSKNNCMQCKMTKKFLDQHGADFEINIDEKPEKIEYVKNLGFSAAPVIEAGNVVFS 63
 +TV+SKNNCMQCKM KK+L +H F EINIDE+PE +E V +GF AAPVI + FS
 30 Sbjct: 2 VTVYSKNNCMQCKMVKKWLSHEHIAFNEINIDEQPEFVEKVIEMGFRAAPVITKDDFAFS 61

Query: 64 GFQPSKLLKEL 73
 GF+PS+L +L
 35 Sbjct: 62 GFRPSELAKL 71

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4475> which encodes the amino acid sequence <SEQ ID 4476>. Analysis of this protein sequence reveals the following:

40 Possible site: 17
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4606(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 56/71 (78%), Positives = 68/71 (94%)

50 Query: 4 ITVFSKNNCMQCKMTKKFLDQHGADFEINIDEKPEKIEYVKNLGFSAAPVIEAGNVVFS 63
 ITV+SKNNCMQCKMTKKFL+QHG +F+EINIDE PEK++YVK+LGF++APVIEA N+VFS
 Sbjct: 13 ITVYSKNNCMQCKMTKKFLQHGQVNFQEIINIDEHPEKVDYVKS LGFTSAPVIEADNLVFS 72

Query: 64 GFQPSKLLKELV 74
 GFQP+KLKEL+
 55 Sbjct: 73 GFQPAKLKELI 83

-1609-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1457

- 5 A DNA sequence (GBSx1543) was identified in *S.agalactiae* <SEQ ID 4477> which encodes the amino acid sequence <SEQ ID 4478>. This protein is predicted to be ribonucleotide reductase subunit R1E (nrdE). Analysis of this protein sequence reveals the following:

Possible site: 49
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3676(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
- 15 The protein has homology with the following sequences in the GENPEPT database.
- >GP:AAD41036 GB:AF112535 ribonucleotide reductase alpha-chain
[Corynebacterium glutamicum]
Identities = 366/701 (52%), Positives = 488/701 (69%), Gaps = 19/701 (2%)
- 20 Query: 23 NGQIPLHKDKKEALTAFKENVQPNKAFDSITDKIAYLLKYDYLEEAFLNKYRPEFIEEL 82
NG+I KD+EA +F ++V N+ F ++ +KI YL++ Y + L+KY +FI++L
Sbjct: 22 NGKIQFEKDREAANQYFLQHVNTVFFHNLQEKIDYLVENKYYDPIVLDKYDFQFIKDL 81
- 25 Query: 83 STKLFDKKFRFKSFMAAYKFYQQYALKTNDGEYYLESIEDRVLFNALYFADGDEELATDL 142
+ + KFRF+SF+ AYK+Y Y LKT DG YLE EDRV AL ADGD LA +L
Sbjct: 82 FKRAYGFKFRFQSLGAYKYYSYTLKTFDGRYLERFEDRVCMVALTLADGDRALAENL 141
- 30 Query: 143 ALEMISQRYQPATPSFLNAGRSRRGELVSCFLIQVTDDMNAIGRSINSALQLSRIGGGVG 202
E++S R+QPATP+FLN+G+++RGE VSCFL+++ D+M +IGRSINSALQLS+ GGGV
Sbjct: 142 VDEIMSGRFQPATPTFLNSGKAQRGEVPVSCFLLRIEDNMESIGRSINSALQLSKRGGGVA 201
- 35 Query: 203 ISLSNLREAGAPIKGFAGAASGVVPMKLFEDSFSYSNQLGQAGVYVLDVFPDIIIS 262
+ LSNLREAGAPIK +SGV+PVMKL ED+FSY+NQLG RQCAG VYL+ HPDI+S
Sbjct: 202 LLSNLREAGAPIKKIENQSSGVIPIVMKLLDAFSYANQLGARQAGAVYLNHHPDILS 261
- 40 Query: 263 FLSTKKNADKVRVKTLSLGITVPDKFYELARNNQEMYLFSPYSIEREYGVFPFSYIDIT 322
FL TK+ENADEK+R+KTLGLG+ +PD +ELA+ N +MYLFSPY +ER YG PF+ + IT
Sbjct: 262 FLDTKKNADKIRIKTSLGVIPIPDITFELAKRNDMYLFSPYDVERIYGKPFADVSI 321
- 45 Query: 323 EKYDELVANPNITKTKINARDETEISKLQESGYPYIINIDTANRTPVDGKIIMSNLC 382
E YDE+V + I KTKINAR ++++Q ESGYPYI+ DT N +NP++G+I SNLC
Sbjct: 322 EHYDEMVDDDRIRKTKINARQFFQTLAEIQFESGYPYIMYEDTVNASNPIEGRITHSNLC 381
- 50 Query: 383 SEILQVQKPSLINDAQEYLEMGTDISCNLGSNTVLNMMTSPDFGKSIKTMTRALTFTVDS 442
SEILQV PS ND Y E+G DISCNLGS NV M SP+F K+I+T R LT V++
Sbjct: 382 SEILQVSTPSEFNDDLTVAEVGEDISCNLGSNLVAMAMDSPNFEKTIETAIRGLTAVSEQ 441
- 55 Query: 443 SNIEAVPTIKNGNAQAHTFGLGAMGLHSYLAKNHIEYGSPESEFTDIYFMLMNYWTLVE 502
++I++VP+I+ GN AH GLG M LH Y + H+ YGS E+++FT+ YF + Y L
Sbjct: 442 TSIDSVPFSIRKGNAAHAIGLQGMNLHGYPGRHMHYGSBEALDFTNAYFAAVLYQCLRA 501
- 60 Query: 503 SNNIARERQTTFFVGFESKSYADGTIFYDKYVSGKFVPQSDKVKSLFA--NHFIPEAKDWEN 560
SN IA ER F FE SKYA G YFD + + F P+SDKVK LFA N P +DW
Sbjct: 502 SNKIATERGERFKNFENS KYATGEYFDDFDANDFAPKSDKVKELFAKSNHTPTIVEDWAA 561
- 65 Query: 561 LRYAVMKDGLYHQNLAVAPNGSISYINDCSASIHPIQRIEERQEKKIGKIYYPANGLA 620
L+ VM+ GL+++N AV P GSISYIN+ ++SIHPI +IE R+E KIG++YYP A +
Sbjct: 562 LKADVMEHGLFNRLQAVPPTGSISYINNSTSSIHPIASKIEIRKEGKIGRVYYPAPHMD 621
- 70 Query: 621 TDTIPYYTSAYDMDMRKVIDVYAAATEHVDQGLSMTLFLRSELPKELYEWKTESKQTTRD 680
D + Y+ AY++ K+ID YA AT++VDQGLS+TLF + TTRD
Sbjct: 622 NDNLEYFEDAYEIGYEKIIDTYAVATKYVDQGLSLTLFFK-----DTATTRD 668

-1610-

Query: 681 LSILRNIAFNKGVSIIYYI--RTFTDDGSEVGANQCESCVI 719
 ++ + YA+ KG+K++YYI R +G+EV + C SC++
 Sbjct: 669 INRAQIYAWRKGIKTLYYIRLRQVALEGTEV--DGCVSCML 707

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4479> which encodes the amino acid sequence <SEQ ID 4480>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4241(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 628/719 (87%), Positives = 682/719 (94%)

Query: 1 MSLKNIGDVSFYRLNNEINRPVNGQIPLHKDKREALTAFFKENVQPNKAFDSITDKIAYL 60
 MSLK++GD+SYFRLNNEINRPVNG+IPLHKDKREAL AF ENV PN+ +F SIT+KI YL
 Sbjct: 1 MSLKDLGDISFYRLNNEINRPVNGKIPLHKDKREALKAFSAENVLPNTMSFTSITEKIEYL 60

Query: 61 LKYDYLEEAFLNKYRPEFIEELSTKLFDKKFRFKSFMAAYKFYQQYALKTNDGEYYLES 120
 + DY+E AF+ KYRPEFI EL + + + FRFKSFMAAYKFYQQYALKTNDGE+YLE++
 Sbjct: 61 ISNDYIESAFIQYRPEFITELDSIIKSENFRRFKSFMAAYKFYQQYALKTNDGEHYLENL 120

Query: 121 EDRVLFNALYFADGDEELATDLALEMISORYQPATPSFLNAGRSRRGELVSCFLIQVTDD 180
 EDRVLFNALYFADG E+LA DLA+EMI+QRYQPATPSFLNAGRSRRGELVSCFLIQVTDD
 Sbjct: 121 EDRVLFNALYFADGQEDLAKDLAVEMINQRYQPATPSFLNAGRSRRGELVSCFLIQVTDD 180

Query: 181 MNAIGRSINSALQLSRIGGGVGISLNLREAGAPIKGAGAAAGVVPVMKLFEDSFSYSN 240
 MN+IGRSINSALQLSRIGGGVGIT+LSNLREAGAPIKG+AGAAAGVVPVMKLFEDSFSYSN
 Sbjct: 181 MNSIGRSINSALQLSRIGGGVGITLSNLREAGAPIKGYAGAAGVVPVMKLFEDSFSYSN 240

Query: 241 QLQQRQGAGVVYLDVFHPDIIISFLSTKKENADEKVRVKTLISLGITVPDKFYELARNNQEM 300
 QLQQRQGAGVVYL+VFHPDII+FLSTKKENADEKVRVKTLISLGITVPDKFYELAR N++M
 Sbjct: 241 QLQQRQGAGVVYLVNLFHPDIIAFLSTKKENADEKVRVKTLISLGITVPDKFYELARKNEDM 300

Query: 301 YLFSPYSIEREYGVPFYSYIDITEKYDELVANPNITKTKINARDLETEISKLOQESGYPYI 360
 YLFSPY++E+EYG+PF+Y+DIT YDELVANP ITKTKI ARDLETEISKLOQESGYPYI
 Sbjct: 301 YLFSPYNVEKEYGIPFNLYDITNMYDELVANPKITKTKIARDLETEISKLOQESGYPYI 360

Query: 361 INIDTANRTNFPVDGKIIMSNNLCSEILQVQKPSLINDAQEYLEMGTDISCNLGSSTNVLMNM 420
 INIDTAN+ NP+DGKIIMSNNLCSEILQVQ PSLINDAQE++EMGTDISCNLGSSTN+LNMNM
 Sbjct: 361 INIDTANKANPIDGKIIMSNNLCSEILQVQTPSLINDAQEFVEMGTDISCNLGSSTNINLMNM 420

Query: 421 TSPDFGKSIKTMTRALTFTVDSSNIEAVPTIKNGNAQAHTFGLGAMGLHSYLAKNHIEYG 480
 TSPDFG+SIKTMTRALTFTVDSS+IEAVPTIK+GN+QAHTFGLGAMGLHSYLA++HIEYG
 Sbjct: 421 TSPDFGRSIKTMTRALTFTVDSSSIEAVPTIKHGNSQAHTFGLGAMGLHSYLAQHIEYG 480

Query: 481 SPESIEFTDIYFMLMNYWTLVESNNIARERQTTFVGFEKSKYADGSTYFDKYVSGKFVPQS 540
 SPESIEFTDIYFML+NYWTLVESNNIARERQTTFVGFE SKYA+G+YFDKYV+G FVP+S
 Sbjct: 481 SPESIEFTDIYFMLLNYWTLVESNNIARERQTTFVGFEKSKYANGSYFDKYVTGHFVPKS 540

Query: 541 DKVKSLEFANHFIEPAKDWNELRYAVMKDGLYHQNRLAVAPNGSISYINDCSASIHPIQR 600
 D VK LF +HFIP+A DWE LR AV KDGLYHQNRLAVAPNGSISYINDCSASIHPIQR
 Sbjct: 541 DLVKDLFKDHFIPQASDWEALRDVAQKDGLYHQNRLAVAPNGSISYINDCSASIHPIQR 600

Query: 601 IEERQEKKIGKIYYPANGLATDTIPYYTSAYDMDMRKVIDVYAAATEHVDQGLSMTLFLR 660
 IEERQEKKIGKIYYPANGI+TDIPYYTSAYDMDMRKVIDVYAAATEHVDQGLS+TLFLR
 Sbjct: 601 IEERQEKKIGKIYYPANGI+TDIPYYTSAYDMDMRKVIDVYAAATEHVDQGLSLTLFLR 660

Query: 661 SELPKELYEWKTESKQTTTRDLSILRNIAFNKGVSIIYYIRFTFTDDGSEVGANQCESCVI 719
 SELP ELYEWKT+SKQTTTRDLSILRNIAFNKG+KSIYYIRFTFTDDG EVGANQCESCVI
 Sbjct: 661 SELPMELYEWKTQSKQTTTRDLSILRNIAFNKGKSIYYIRFTFTDDGEEVGANQCESCVI 719

-1611-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1458

- 5 A DNA sequence (GBSx1544) was identified in *S.agalactiae* <SEQ ID 4481> which encodes the amino acid sequence <SEQ ID 4482>. This protein is predicted to be ribonucleotide reductase subunit R2F (nrdB). Analysis of this protein sequence reveals the following:

```

Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4583(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

15 A related GBS nucleic acid sequence <SEQ ID 9753> which encodes amino acid sequence <SEQ ID 9754> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC14561 GB:AF050168 ribonucleoside diphosphate reductase small
subunit [Corynebacterium ammoniagenes]
Identities = 166/313 (53%), Positives = 215/313 (68%), Gaps = 1/313 (0%)

Query: 10  EAINWNEIEDVIDKSTWEKLTEQFWLDTRIPLSNDLDDWRKLSAQEKDLVGKVFGGLTLL 69
      +AINWN I D D  W++LT FWL +IP+SND+ W K++ QE+ +VF GLTLL
Sbjct: 17  KAINWNVIPDEKDLVWDRLTGNFWLPEKIPVSNDIQSWNKMTPOEQLATMRVFTGLTLL 76

Query: 70  DTMQSETGVEAIRADVTRPHEEAVLNNIQFMESVHAKSYSSIFSTLNTKSEIEEIFEWTN 129
      DT+Q  G ++ DV T HEE V NI FMESVHAKSYS+IF TL + +I E F W+
Sbjct: 77  DTIQGTVGAIISLLPDVETMHEEGVYTNI AFMESVHAKSYSNIFMTLASTPQINEAFRWSE 136

Query: 130  NNEFLQEKARIINDIYANGNALQKKVASTYLETFLFYSGFFTPLYLGNNKLANVAEIIK 189
      NE LQ KA+II Y + L+KKVAST LE+FLFYSGF+ P+Y KL N A+II+
Sbjct: 137  ENENLQRKAKIIMSYYNGDDPLKKKVASTLLESFIFYSGFYLPMYLSRAKLNTADIIR 196

Query: 190  LIIRDESVMHGTYYIGYKQFQGFNPEDEQENFRDWMYDLYQLYENEEKYTKTLYDGVGW 249
      LIIRDESVMHGTYYIGYK+Q G +L E EQE ++ + +DL+Y LYENE +YT+ +YD +GW
Sbjct: 197  LIIRDESVMHGTYYIGYKQGVKKLSEAEQEYKAYTFDLMYDLYENEIEYTEDIYDDL GW 256

Query: 250  TEEVMTFLRYNANKALMNLGQDPLFPDPTANDVNPVIMNGIS-TGTSNHDFFSQVGN GYLL 308
      TE+V FLRYNANKAL NLG + LFP V+P +++ +S NHDFFS G+ Y++
Sbjct: 257  TEDVKRFLRYNANKALNNLGYEGLFPTDETQVSPAILSSLSPNADENHDFFSGSGSSYVI 316

Query: 309  GSVEAMHDDDDYNY 321
      G E DDD+++
Sbjct: 317  GKAEDTTDDDDWDF 329

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4483> which encodes the amino acid sequence <SEQ ID 4484>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4583(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

55 An alignment of the GAS and GBS proteins is shown below.

-1612-

Identities = 315/319 (98%), Positives = 316/319 (98%)

Query: 5 MTTYEAINWNEIEDVIDKSTWEKLTQFWLDTRIPLSNDLDDWRKLSAQEKDLVGKVFG 64
 Sbjct: 1 MTTYEAINWNEIEDVIDKSTWEKLTQFWLDTRIPLSNDLDDWRKLSAQEKDLVGKVFG 60

Query: 65 GLTLDDTMQSETGVEAIRADVTPHEEAVLNNIQFMESVHAKSYSSIFSTLNTKSEIEEI 124
 Sbjct: 61 GLTLDDTMQSETGVEAIRADVTPHEEAVLNNIQFMESVHAKSYSSIFSTLNTKSEIEEI 120

Query: 125 FEWTNNNEFLQEKARIINDIYANGNALQKKVASTYLETFLFYSGFFTPLYYLGNKLANV 184
 Sbjct: 121 FEWTNNNEFLQEKARIINDIYANG+ALQKKVASTYLETFLFYSGFFTPLYYLGNKLANV 180

Query: 185 AEIIKLIIRDES VHGTIYIGYKFQLGFNELPEDEQENFRDWMYD LLYQLYENEEKYTKTLY 244
 Sbjct: 181 AEIIKLIIRDES VHGTIYIGYKFQLGFNELPEDEQENFRDWMYD LLYQLYENEEKYTKTLY 240

Query: 245 DGVGWTEEVMTFLRYNANKALMNLGQDPLFPDTANDVNPIVMNGISTGT SNHDFFSQVGN 304
 Sbjct: 241 DGVGWTEEVMTFLRYNANKALMNLGQDPLFPDTANDVNPIVMNGISTGT SNHDFFSQVGN 300

Query: 305 GYLLGSVEAMHDDDDYNYGL 323
 GYLLGSVEAM DDDYNYGL
 Sbjct: 301 GYLLGSVEAMSDDDYNYGL 319

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1459

A DNA sequence (GBSx1545) was identified in *S. agalactiae* <SEQ ID 4485> which encodes the amino acid sequence <SEQ ID 4486>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.27 Transmembrane 50 - 66 (50 - 66)

----- Final Results -----
 bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1460

A DNA sequence (GBSx1546) was identified in *S. agalactiae* <SEQ ID 4487> which encodes the amino acid sequence <SEQ ID 4488>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -14.38 Transmembrane 176 - 192 (168 - 201)
 INTEGRAL Likelihood = -4.57 Transmembrane 25 - 41 (22 - 42)
 INTEGRAL Likelihood = -3.88 Transmembrane 94 - 110 (94 - 112)
 INTEGRAL Likelihood = -1.49 Transmembrane 70 - 86 (70 - 86)
 INTEGRAL Likelihood = -1.01 Transmembrane 128 - 144 (128 - 144)

----- Final Results -----

-1613-

bacterial membrane --- Certainty=0.6753(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9751> which encodes amino acid sequence <SEQ ID 9752> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15077 GB:Z99119 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 55/184 (29%), Positives = 98/184 (52%), Gaps = 4/184 (2%)

10 Query: 16 MSKNNTTCLITAIFAALAMALSMIP---DFASWFTPSFGAIPILILFALRRGTKYGLF 71
 M+++ LIE AI A A+ L ++ + S IP+ L + R G K GL
 Sbjct: 1 MNQSKQLVRLIEIAIMTAAAVILDIVSGMFLSMPQGGSVSIMMIPIFLISFRWGVKAGLT 60

15 Query: 72 AGLIWGLLHFVLSKVYVLSLSQVFIEYILAFISMGLAGVFSAKFKDALSSSSSKTKALSLA 131
 GL+ GL+ + ++ Q+ ++YI+AF ++G++G F++ + A S +K K +
 Sbjct: 61 TGLLTGLVQIAIGNLFAQHVPVQLLLDYIVAFAAIGISGCFASSVRKAAVSKTKGKLIVSV 120

20 Query: 132 LSGAILATLVRYVWHYIAGVIFWASYAPKGMSATLYSLSVNGTAGLLTLFFVVISIIILV 191
 +S + +L+RY H I+G +F+ S+APKG +YSL+ N T + + I + +L
 Sbjct: 121 VSAVFIGSLRLRYAAHVISGAVFFGFSAPKGTTPVWIYSLTYNATYMVPSFIICAIVLCLLF 180

Query: 192 ISYP 195
 ++ P
 25 Sbjct: 181 MTAP 184

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4489> which encodes the amino acid sequence <SEQ ID 4490>. Analysis of this protein sequence reveals the following:

Possible site: 20
 30 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -9.34 Transmembrane 162 - 178 (156 - 183)
 INTEGRAL Likelihood = -9.34 Transmembrane 110 - 126 (107 - 130)
 INTEGRAL Likelihood = -1.22 Transmembrane 55 - 71 (55 - 71)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 40 The protein has homology with the following sequences in the databases:

>GP:CAB15077 GB:Z99119 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 55/189 (29%), Positives = 100/189 (52%), Gaps = 10/189 (5%)

45 Query: 1 MSPNTNVKYLIEAAIFAALAMTSLFIPDFAGWF--SPSYGAIALV-----IFSLRRGLKY 53
 M+ + + LIE AI A A+ L + +G F P G+++++ + S R G+K
 Sbjct: 1 MNQSKQLVRLIEIAIMTAAAVILDIV---SGMFLSMPQGGSVSIMMIPIFLISFRWGVKA 57

Query: 54 GMLAGLIWGLLHFVLGKVVYLSMSQVFIEYILAFISMGLAGSFSDSLKTLRRQQTFFAV 113
 G+ GL+ GL+ +G ++ Q+ ++YI+AF ++G++G F+ S+ K + +
 50 Sbjct: 58 GLTTGLLTGLVQIAIGNLFAQHVPVQLLLDYIVAFAAIGISGCFASSVRKAAVSKTKGKLI 117

Query: 114 FLAIMASLLAVTVRYLWHFLAGIIFWGSYAPKGMSAVWYSFVNGTAGVLTFLITCLALM 173
 + A + +RY H ++G +F+GS+APKG YS + N T V +F+I + L
 Sbjct: 118 VSVSAVFIGSLRLRYAAHVISGAVFFGFSAPKGTTPVWIYSLTYNATYMVPSFIICAIVLC 177

55 Query: 174 IALPIHPQL 182
 + P+L
 Sbjct: 178 LLFMTAPRL 186

- 60 An alignment of the GAS and GBS proteins is shown below.

Identities = 116/186 (62%), Positives = 138/186 (73%)

-1614-

Query: 16 MSKNNNTTCLIIETAIFAALAMALSMIPDFASWFTSPFGAIPILIFALRRGTKYGLFAGLI 75
 MS N N LIE AIFAALAM LS IPDFA WF+PS+GAI L++F+LRRG KYG+ AGLI
 Sbjct: 1 MSPNTNVKYLIEAIFAALAMTSLFIPDFAGWFSYGAIALVIFSLRRGLKYGMLAGLI 60

Query: 76 WGLLHFVLSKVYYLSLSQVFIEYILAFISMGLAGVFSKFKDALSSSSSKTKALSLALSGA 135
 WGLLHFVL KVVYLS+SQVFIEYILAF SMGLAG FS L A+ LA+ +
 Sbjct: 61 WGLLHFVLGKVYYLSMSQVFIEYILAFISMGLAGSFSDSLIKTLRRQQTFFAVFLAIMAS 120

Query: 136 ILATLVRYVWHYIAGVIFWASYAPKGM SATLYSLSVNGTAGLLTLFFVVISIILVISYP 195
 +LA VRY+WH++AG+IFW SYAPKGM SA YS SVNGTAG+LT ++++I + +P
 Sbjct: 121 LLAVTVRYLWHFLAGIIFWGSYAPKGM SAVWYSFVNGTAGVLTFLTCLALMIALPIHP 180

Query: 196 SFFFLPK 201
 F PK
 Sbjct: 181 QLFDPK 186

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 1461

A DNA sequence (GBSx1547) was identified in *S.agalactiae* <SEQ ID 4491> which encodes the amino acid sequence <SEQ ID 4492>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

25	INTEGRAL	Likelihood = -7.43	Transmembrane	206 - 222 (199 - 223)
	INTEGRAL	Likelihood = -6.64	Transmembrane	24 - 40 (19 - 42)
	INTEGRAL	Likelihood = -6.58	Transmembrane	61 - 77 (51 - 78)
	INTEGRAL	Likelihood = -6.58	Transmembrane	134 - 150 (132 - 154)
	INTEGRAL	Likelihood = -4.62	Transmembrane	226 - 242 (224 - 245)
30	INTEGRAL	Likelihood = -3.72	Transmembrane	107 - 123 (106 - 125)

----- Final Results -----
 bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9749> which encodes amino acid sequence <SEQ ID 9750> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4493> which encodes the amino acid sequence <SEQ ID 4494>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

45	INTEGRAL	Likelihood = -10.46	Transmembrane	134 - 150 (131 - 159)
	INTEGRAL	Likelihood = -7.59	Transmembrane	107 - 123 (103 - 128)
	INTEGRAL	Likelihood = -7.48	Transmembrane	225 - 241 (213 - 248)
	INTEGRAL	Likelihood = -7.22	Transmembrane	205 - 221 (199 - 224)
	INTEGRAL	Likelihood = -3.56	Transmembrane	50 - 66 (50 - 73)
50	INTEGRAL	Likelihood = -1.28	Transmembrane	16 - 32 (16 - 33)

----- Final Results -----
 bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

-1615-

An alignment of the GAS and GBS proteins is shown below.

Identities = 82/253 (32%), Positives = 149/253 (58%), Gaps = 5/253 (1%)

```

5  Query: 6  IKQSDTTFVRIIKSLLIGGFIGAILGSGVGFIIIF--GQDKYLSEI--NIVQYFLWVSRI 61
   +K+  +F+R++K L+  G I+G +  F+ +  G+  +L+ +  +++  + ++R+
Sbjct: 1  MKKKKNSFLRLLKMSLLSSLAGGIIGMVGAFGLGYHGGRLDHLTFLKDDVINLIILLNRL 60

10 Query: 62  VVIITALFSLIYLYQIQYQKVFFNVDESQ--SEETIRQINLRHSYGMTFVSISIVLSIVN 120
   VV+  S ++L Q++K V+  ++E  SE YRQ+N +H+Y M  ++++ +LS+ N
Sbjct: 61  VVVTDLTLSFVFLTQLKKETAVYNTIBEDDISENGYRQLNKKHAYTMLLIIVASILSMCN 120

15 Query: 121  TLFNFKLNIFFDDSVTLVIPIYDLSLLFVLLGLHLYFLKVYRNIRGIKMTVAPTLLKELKNN 180
   L  L  L IP+ D+ LL +++  +K Y  IRG  +  P LKELK+N
Sbjct: 121  VLLGLTLTND SQHAMLAIPLLDILLMLVFPQALAMKRYNAIRGTDVPYFPNLKELKHN 180

20 Query: 181  VLQDLAEAELESNYKMCFDIVMNLSGFIFPTIYFVLFFFISFVFQKVEIVAIITTSIHIYI 240
   ++ LDEAEL++ +K  F+ V++L+G I P++Y +LFF+  +VE+ AI++  I +Y+
Sbjct: 181  IMALDEAELQAYHKTSFESVLSLNGVLIIPSLYVILFFVYLFVGQVELTAILVLVLQLYL 240

20 Query: 241  LIKSLKAARHFYR 253
   L+KS  R FYR
Sbjct: 241  LVKSATMTRQFYR 253

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1462

A DNA sequence (GBSx1548) was identified in *S.agalactiae* <SEQ ID 4495> which encodes the amino acid sequence <SEQ ID 4496>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 57
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
           bacterial cytoplasm --- Certainty=0.5172(Affirmative) < succ>
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1463

A DNA sequence (GBSx1549) was identified in *S.agalactiae* <SEQ ID 4497> which encodes the amino acid sequence <SEQ ID 4498>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 26
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
           bacterial cytoplasm --- Certainty=0.2059(Affirmative) < succ>
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC76650 GB:AE000440
   UDP-D-glucose: (galactosyl) lipopolysaccharide

```

-1616-

glucosyltransferase [Escherichia coli K12]
 Identities = 70/256 (27%), Positives = 121/256 (46%), Gaps = 14/256 (5%)

5 Query: 1 MNLLFSIDDMYVDHFKVMLYSLVRQTKNRKLEIYVLQKT----LLKRHTELIQYTONLEV 56
 +N+ + +D Y+D V + S+V ++ L+ Y++ ++ +L + Q
 Sbjct: 28 LNVAYGVDANYLDGVGVSITSIVLNNRHINLDFYIIADVYNDGFFQKIAKLAEQNQLRIT 87

10 Query: 57 GYHPPIIVGTEVFAQAFTTD RYPTIYRLLAHKFLPETLDRIYLDADMLCLNDFSSLYD 116
 Y + T+ P T + +Y+RL A + L TLDR+LYLDAD++C D S L
 Sbjct: 88 LYR---INTDKLQCLPCTQVWSRAMYFRLFAFQQLGLTLDRLLYLDADVCKGDISQLLH 144

15 Query: 117 MELGDQLYAAASHNTDGKFLDYVNKLRLKNVELESSYFNTGVLLMNLPAIRKVVHQQTIL 176
 + L A A+ D + + RL + EL YFN+GV+ ++L + L
 Sbjct: 145 LGING---AVAAVVKDVEPMQEKAVSRSLSPPELLGQYFNSGVVYLDLKKWADAKLTEKAL 201

20 Query: 177 DYIMQNRGRLLLPDQDILNGLYANLVKPIPDEIYNYDARYSLIYQLKSRNEWDLEWVINH 236
 +M PDQD++N L + +P E Y+ Y++ +LK + + +I
 Sbjct: 202 SIILMSKDNVYKYPDQDVMNVLLKGMTLFLPRE---YNTIYTIKSELKDKTHQNYKKLITE 258

20 Query: 237 -TVFLHFAGRDKPWK 251
 T+ +H+ G KPW K
 Sbjct: 259 STILLIHYTGATKPWHK 274

No corresponding DNA sequence was identified in *S.pyogenes*.

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1464

A DNA sequence (GBSx1550) was identified in *S.agalactiae* <SEQ ID 4499> which encodes the amino acid sequence <SEQ ID 4500>. Analysis of this protein sequence reveals the following:

30 Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.1406(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1465

A DNA sequence (GBSx1551) was identified in *S.agalactiae* <SEQ ID 4501> which encodes the amino acid sequence <SEQ ID 4502>. Analysis of this protein sequence reveals the following:

45 Possible site: 54
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -10.72	Transmembrane	7 - 23 (1 - 28)
INTEGRAL	Likelihood = -4.30	Transmembrane	222 - 238 (216 - 238)
INTEGRAL	Likelihood = -3.66	Transmembrane	151 - 167 (140 - 170)
50 INTEGRAL	Likelihood = -3.50	Transmembrane	35 - 51 (34 - 58)
INTEGRAL	Likelihood = -3.35	Transmembrane	71 - 87 (69 - 88)
INTEGRAL	Likelihood = -3.29	Transmembrane	113 - 129 (113 - 132)
INTEGRAL	Likelihood = -2.81	Transmembrane	170 - 186 (168 - 190)
INTEGRAL	Likelihood = -2.71	Transmembrane	198 - 214 (197 - 217)

-1617-

----- Final Results -----

5 bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:BA07774 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 84/242 (34%), Positives = 147/242 (60%), Gaps = 16/242 (6%)

Query: 1 MVGLGTVINVILIIVGGFVGLPLKNFLKESLQKSLMQAMGVAVLFISISGVLEKMMLEK 60
 MV +GTV+N I++ +GL +KN + E ++ +LMQA+G+A++ + + KM L +
 Sbjct: 1 MVLIGTVVNGAAIVIAALIGLLVKN-IPERVKTTLMQAIGLAIVLLGV-----KMGLQTE 54

15 Query: 61 SHLISNHTNMIIITLALGTVLGEALLSDSYIDKFGNYLKQKTGSGNDIKFVEAFVTSTCT 120
 LI +I +L +G V+GE+++L+ +D G +++ K G D AFVT+T
 Sbjct: 55 QFLI-----VICSLVIGGVIGEMINLEKRLDHLGRWIESKVGKKDGSIAATAFVTTTLI 108

20 Query: 121 VCIGAMAVVSGSIQDGIADHSILFAKGMMLDIIITAMTVSLGKGFALFSAIPVALLQGSIT 180
 +GAMAV+G++ G+ DHS+L K +LD + + T +LG G LFSA+PV L QGS+
 Sbjct: 109 YVVGAMAVLGALDSGLRGDHSVLLTKALLDGFLAILFTSLGIGVLFSAIPVVLYQGSIA 168

25 Query: 181 IVAF----FMGSLNPSSLDYLNVLGNMLIFCVGVNLLFNLNIKVINMLPAIILAILWGS 236
 + A ++ + L S + ++ G ++I +G+NLL +NI+V N+LP++++ + +
 Sbjct: 169 LFASQIDQVPTALMDSFITEMSATGGVMIVAIGLNLNVLNIRVANLLPSLVIVAVLVLT 228

Query: 237 FI 238
 F+
 Sbjct: 229 FV 230

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1466

35 A DNA sequence (GBSx1552) was identified in *S.agalactiae* <SEQ ID 4503> which encodes the amino acid sequence <SEQ ID 4504>. This protein is predicted to be alanyl-tRNA synthetase (alaS). Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence
 40 INTEGRAL Likelihood = -4.41 Transmembrane 805 - 821 (804 - 822)

----- Final Results -----
 bacterial membrane --- Certainty=0.2763(Affirmative) < succ>
 45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:BA04986 GB:AP001511 alanyl-tRNA synthetase [Bacillus halodurans]
 Identities = 482/885 (54%), Positives = 618/885 (69%), Gaps = 27/885 (3%)

Query: 1 MKELSSAQIRQMWLDFWKSCKHSVEPSANLVPVNDPTLLWINSVATLKKYFDGSPVIPEN 60
 MK L+SAQ+RQM+LDF+K KGH VEPSA+LVP +DP+LLWINSVATLKKYFDG VIPEN
 Sbjct: 1 MKYLTSAQVRQMFLDFFKEKGHDVEPSASLVPHDDPSLLWINSVATLKKYFDGRVIPEN 60

55 Query: 61 PRITNAQKSIRTNDIENVGKTARHHTMFEMLGNFSIGDYFRDEAIEWGFELLTSPEWFDF 120
 PRITNAQKSIRTNDIENVGKTARHHT FEMLGNFSIGDYF++EAIEW +E LTS +W F
 Sbjct: 61 PRITNAQKSIRTNDIENVGKTARHHTFFEMLGNFSIGDYFKEEAIEWAWEFLTSEKWIGF 120

Query: 121 PKDKLYMTYYPDDKDSYNRWIA-CGVEPSHLVPIEDNFWIEGAGPSGPDTEIFFDRGEDF 179

-1618-

K+KL +T +P+D ++Y+ W G+ ++ +E NFW+IG GPSGP+TEIF+DRG ++
 Sbjct: 121 DKEKLSVTVHPEDDEAYSYWKEKIGIPEERIIRLEGNFWDIGEGPSGPNTTEIFYDRGPBY 180
 Query: 180 -----DPENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGL 234
 DPE L ENDRY+E+WN+V SQFN +P Y LP KNIDTG GL
 Sbjct: 181 GDQNPDE-----LYPGGENDRYLEVWNLVFSQFNHNPD---GSYTPLPKKNIDTGMGL 231
 Query: 235 ERLAAVMQGAKTNFETDLFMPPIIREVEKLSGKTYDPDGD-NMSFKVIADHIRALSFAIGD 293
 ER+ +V+Q TNFETDLFMPPIR EK+SG Y + ++SFKVIADHIR ++FAIGD
 Sbjct: 232 ERMVSVIQNVPTNFETDLFMPPIIRATEKISGTEYGSHEADVSPFKVIADHIRTVTFAIGD 291
 Query: 294 GALPGNEGRGYVLRRLRRRAVMHGRRLGINETFLYKLVPTVGQIMESYYPEVLEKRDFFIE 353
 GALP NEGRGYVLRRLRRRAV + +++GL+ F+Y+LVP VG IM +YPEV EK FI+
 Sbjct: 292 GALPSNEGRGYVLRRLRRRAVRYAKQIGIDRPFMYELVPVVGIMVDFYPEVKEKAAPFIQ 351
 Query: 354 KIVKREEETFARTIDAGSGHLDSLLAQLKAEGKDTLEGKDIFKLYDTYGFVPELTELAE 413
 K+VK EEE F T++ G L+ ++ + K+EG T+ G D+F+LYDTYGFVPV+LTEE E
 Sbjct: 352 KVKTEEERFHETLNEGLSILEKVIDKAKSEGASTISGSDVFRLYDTYGFVPDLTEEYVE 411
 Query: 414 DAGYKIDHEGFSAMKEQQDRARAADVKGSGMGMQNETLAGIVEESRF-EYDYSLESSL 472
 + G ++D +GF++ M+ Q++RAR A + GSM +Q+E L I +S F Y S E+++
 Sbjct: 412 EQQLQVLDLGFEAEMERQREARTARQQAGSMQVQDEVLGQITVDSTFIGYKQLSTETTI 471
 Query: 473 SVIIADNTERTEAVSEGO-ALLVFAQTFFYAEMGGQVADHGVKNDKGDVAVVVDVQKAP 531
 I+ D + V GQ A ++ +TPFYAE GGQVAD G+I+ G V V DVQKAP
 Sbjct: 472 ETIVLDKTVADYVGAGQEAQVILKETPFYAESGGQVADKGIIRGANGFAV--VSDVQKAP 529
 Query: 532 NGQLHTVNVL-ASLSVGTNYTLEINKERRLAVEKNHTATHLLHAALHNVIGEHAQAGS 590
 NGQ LHTV V +L V + + R + KNHTATHLLH AL +V+GEH QAGS
 Sbjct: 530 NGQLHTVIVKEGTLQVNDQVQAIIVEETERSGIVKNHTATHLLHRAKLDVLGEHVNQAGS 589
 Query: 591 LNNEEFRLRDFTHFEAVSNEELRHIEQEVNEQIWNDLTITTTTETDVTAKEMGAMALFGE 650
 L EE LRFDF+HF V++EE IE+ VNE+IW + + + ++ AK +GAMALFGE
 Sbjct: 590 LVSEERLRLRDFSHFGQVTDDEEKEKIERIVNEKIWQAIKVNISTKTLDCAKAGAMALFGE 649
 Query: 651 KYGKVVVVQIGNYSVELCGGTHLNSSEIGLFKIVKEEGIGSGTRRIIAVTGRQAFAY 710
 KYG +VRVV++G+YS+ELCGG H+ N+SEIGLFKIV E GIG+G RRI AVTG++AF
 Sbjct: 650 KYGDIVRVVEVDYSIELCGGCHVTNTSEIGLFKIVSESGIGAGVRRIEAVTGKEAFLEFM 709
 Query: 711 RNQEDALKRIAATVKAPQLKDAKQALSDSLRDLQKENVELKEAAAAAGDVFKDIQ 770
 Q D LKE AATVKA +KD +V+AL +R+LQ+EN L K AG + ++Q
 Sbjct: 710 AKQLDLLKETAATVKAKNVKDVPRVEALQQQIRELQRENESLNKLGNMAGSLVNEVQ 769
 Query: 771 EAKGVRFIASQVDVADAGALRTFADNWKQKDYSVVLVVAIGEKNVNLVASKTKDV--- 827
 + +GV +A + AD LR+ D KQ+ S V+VL A KVN+ VA TKD+
 Sbjct: 770 KIEGVPVLAKAISGADMDGLRSIVDKLKQEIIPSVVIVLGTASEGKVNIVAGVTKDLINK 828
 Query: 828 --HAGNMIKGLAPIVAGRGGKPDMAAGGSDASKIAELLAABAE 870
 HAG ++K +A G GGG+PDMA AGG K+ + L+ V E
 Sbjct: 829 GYHAGKLVKEVATRCGGGGGGRPDMAQAGGKQPEKLQDALS FVYE 873

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4505> which encodes the amino acid sequence <SEQ ID 4506>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.41 Transmembrane 805 - 821 (804 - 822)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2763(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 862/870 (99%), Positives = 864/870 (99%)

-1619-

Query: 1 MKELSSAQIRQMWLD FWKSKGHSEPSANLVPVNDPTLLWINSGVATLKKYFDG SVIPEN 60
 Sbjct: 1 MKELSSAQIRQMWLD FWKSKGHSEPSANLVPVNDPTLLWINSGVATLKKYFDG SVIPEN 60

5 Query: 61 PRITNAQKSIRINDIENVGKTARHHTMFEMLGNFSIGDYFRDEAIEWGFELLTSPWFDF 120
 PRITNAQKSIRINDIENVGKTARHHTMFEMLGNFSIGDYFRDEAIEWGFELLTSP+WFDF
 Sbjct: 61 PRITNAQKSIRINDIENVGKTARHHTMFEMLGNFSIGDYFRDEAIEWGFELLTSPWFDF 120

10 Query: 121 PKDKLYMTYYPDDKDSYNRWIACGVEPSHLVPIEDNFWEIGAGPSGPDTEIFFDRGEDFD 180
 PKDKLYMTYYPDDKDSYNRWIACGVEPSHLVPIEDNFWEIGAGPSGPDTEIFFDRGEDFD
 Sbjct: 121 PKDKLYMTYYPDDKDSYNRWIACGVEPSHLVPIEDNFWEIGAGPSGPDTEIFFDRGEDFD 180

15 Query: 181 PENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGLERLAAV 240
 PENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGLERLAAV
 Sbjct: 181 PENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGLERLAAV 240

20 Query: 241 MQGAKTNFETDLEMPPIIREVEKLSGKTYDPDGDNMSFKVIADHIRALSFAIGDGALPGNE 300
 MQGAKTNFETDLEMPPIIREVEKLSGKTYDPDGDNMSFKVIADHIRALSFAIGDGALPGNE
 Sbjct: 241 MQGAKTNFETDLEMPPIIREVEKLSGKTYDPDGDNMSFKVIADHIRALSFAIGDGALPGNE 300

25 Query: 301 GRGYVLRLLRRAVMHGRRLGINETFLYKLVPTVGQIMESYYPEVLEKRD FIEKIVKREE 360
 GRGYVLRLLRRAVMHGRRLGINETFLYKLVPTVGQIMESYYPEVLEKRD FIEKIVKREE
 Sbjct: 301 GRGYVLRLLRRAVMHGRRLGINETFLYKLVPTVGQIMESYYPEVLEKRD FIEKIVKREE 360

30 Query: 361 ETFARTIDAGSGHLSLLAQLKAEKDTLEGKDIKLYDTYGFPPVELTEELADAGYKID 420
 ETFARTIDAGSGHLSLLAQLKAEKDTLEGKDIKLYDTYGFPPVELTEELADAGYKID
 Sbjct: 361 ETFARTIDAGSGHLSLLAQLKAEKDTLEGKDIKLYDTYGFPPVELTEELADAGYKID 420

35 Query: 421 HEGFKSAMKEQQDRARA AVVKGSGMGMQNETLAGIVEESRFEYDTYSLESSLSVIIADNE 480
 HEGFKSAMKEQQDRARA AVVKGSGMGMQNETLAGIVEESRFEYDTYSLESSLSVIIADNE
 Sbjct: 421 HEGFKSAMKEQQDRARA AVVKGSGMGMQNETLAGIVEESRFEYDTYSLESSLSVIIADNE 480

40 Query: 481 RTEAVSEGQALLVFAQTFFYAEMGGQVADHGVIKNDKGD T VAEVVDVQKAPNGQPLHTVN 540
 RTEAVSEGQALLVFAQTFFYAEMGGQVAD G IKNDKGD T VAEVVDVQKAPNGQPLHTVN
 Sbjct: 481 RTEAVSEGQALLVFAQTFFYAEMGGQVADTGRIKNDKGD T VAEVVDVQKAPNGQPLHTVN 540

45 Query: 541 VLASLSVGTNYTLEINKERRLA VEKNHTATHLLHAALHN VIGEHA TQAGSLNEEEFLRFD 600
 VLASLSVGTNYTLEINKERRLA VEKNHTATHLLHAALHN VIGEHA TQAGSLNEEEFLRFD
 Sbjct: 541 VLASLSVGTNYTLEINKERRLA VEKNHTATHLLHAALHN VIGEHA TQAGSLNEEEFLRFD 600

50 Query: 601 FTHFEAVSNEELRHIEQEVNEQIWNDLTITTTETD VETAKEMGAMALFGEKYGKVVRVQ 660
 FTHFEAVSNEELRHIEQEVNEQIWN LTITTTETD VETAKEMGAMALFGEKYGKVVRVQ
 Sbjct: 601 FTHFEAVSNEELRHIEQEVNEQIWNALTITTTETD VETAKEMGAMALFGEKYGKVVRVQ 660

55 Query: 661 IGNYSVELCGGTHLNNSEIGLFKIVKEEGIGSGTRRI IAVTGRQAFEAYRNQEDALKEI 720
 IGNYSVELCGGTHLNNSEIGLFKIVKEEGIGSGTRRI IAVTGRQAFEAYRNQEDALKEI
 Sbjct: 661 IGNYSVELCGGTHLNNSEIGLFKIVKEEGIGSGTRRI IAVTGRQAFEAYRNQEDALKEI 720

60 Query: 721 AATVKAPQLKDAAAKVQALSDSLRDLQKENVELKEKAAAAAGDVFKDIQEAKGVRFIAS 780
 AATVKAPQLKDAAAKVQALSDSLRDLQKEN ELKEKAAAAAGDVFKD+QEAKGVRFIAS
 Sbjct: 721 AATVKAPQLKDAAAKVQALSDSLRDLQKENAELKEKAAAAAGDVFKDVQEAKGVRFIAS 780

Query: 781 QVDVADAGALRTFADNWKQKDYS DVLVLVAAIGEKVNVLVASKTKDVHAGNMIK LAPIV 840
 QVDVADAGALRTFADNWKQKDYS DVLVLVAAIGEKVNVLVASKTKDVHAGNMIK LAPIV
 Sbjct: 781 QVDVADAGALRTFADNWKQKDYS DVLVLVAAIGEKVNVLVASKTKDVHAGNMIK LAPIV 840

Query: 841 AGRGGGKPDMA MAGGSDASKIAELLA AAVE 870
 AGRGGGKPDMA MAGGSDASKIAELLA AAVE
 Sbjct: 841 AGRGGGKPDMA MAGGSDASKIAELLA AAVE 870

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1620-

Example 1467

A DNA sequence (GBSx1553) was identified in *S.agalactiae* <SEQ ID 4507> which encodes the amino acid sequence <SEQ ID 4508>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 56
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2974(Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9747> which encodes amino acid sequence <SEQ ID 9748> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAB15920 GB:Z99123 yxjI [Bacillus subtilis]
      Identities = 42/144 (29%), Positives = 73/144 (50%), Gaps = 2/144 (1%)

      Query: 17  IKEKMFSLGGKFTITDLTGLPCYHVEGSLFPLPKTFKVFDEEHLISQIEKKVLSFLPKF 76
      +K+KMFS   F I D       + VEG F L + ++ D       + IE+K++S LP++
20  Sb|ct: 6    MKQKMFSFKDAFHIYDRDEQETFKVEGRFFSLGDSLQMTDSSGKTLVSIQKLSLLPRY 65

      Query: 77  NVTLANGNHFTIKKDFSFLKPHYTTIEDLDMEVKGNFWDMDFQLLKDNQVIANISQQWFRM 136
      +++      + K +F KP + I L+ E+ G+ W +FQL      V ++S++W
25  Sb|ct: 66   EISIGGKTVCEVIKKVTFFSKPKFVISGLNWEIDGDLWRDEFQLTDGENVRMSVSKKWLW 125

      Query: 137 TSTYQVEVYSETYNDLTISLVIAI 160
      +Y +++   E   D+ I   IAI
30  Sb|ct: 126 GDSYHLQIAYE--EDVLICTAIAI 147

```

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1468

35 A DNA sequence (GBSx1554) was identified in *S.agalactiae* <SEQ ID 4509> which encodes the amino acid sequence <SEQ ID 4510>. Analysis of this protein sequence reveals the following:

```

      Possible site: 30
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
40     bacterial cytoplasm --- Certainty=0.3833(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

45  >GP:BAA36674 GB:AB016282 ORF17 [bacteriophage phi-105]
      Identities = 45/133 (33%), Positives = 74/133 (54%), Gaps = 5/133 (3%)

      Query: 2   KYTYLALFEVDKENGGINISFFDFHGAFSEADSLNEAIFNAREVLEIYTIMFEDEGKEFP 61
      +Y Y ALF+ D + G ++FPD G + +S EA+ A+E + ++ FE +G P
50  Sb|ct: 5     RYIYPALFDYDDD--GITVTFFDLPGCITFCGNSGGEALTMKEAMALHLYGFEQDGDIIIP 62

      Query: 62  KASSFKALASNLASDEDVIQAIQSVDTLVRERERSKIVNKTVTLPWSLVEVGKENVNFS 121
      +A+ K +   A +   + I       R       + V KT+T+P W+ ++ KE+KVN+S
55  Sb|ct: 63   EATPSKEIK---AEESQSVVLIETWMPFRHDMENAAVKKTLTIPIRWMDIAKEHKVNYS 119

```

-1621-

Query: 122 QLLQKAIREELQV 134
 QLLQ+AI+E L +
 Sbjct: 120 QLLQEAIKEHLGI 132

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1469

- 10 A DNA sequence (GBSx1555) was identified in *S.agalactiae* <SEQ ID 4511> which encodes the amino acid sequence <SEQ ID 4512>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1484(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 20 >GP:BAA25696 GB:AB010712 NADH oxidase/alkyl hydroperoxidase
 reductase [Streptococcus mutans]
 Identities = 383/509 (75%), Positives = 441/509 (86%)
- 25 Query: 1 MVLDEIKAKLAQYLDLLES DIVLQADLGDNDNSQKVDFLDEIVAMSDRISLESTHLKR 60
 M LD EIK QL QYL LLES+IVLQA L D+ NSQKVK+FL EIVAMS ISLE L R
 Sbjct: 1 MALDAEIKELQYLDLLES DIVLQADLGDNDNSQKVDFLDEIVAMSPMISLEEKELPR 60
- 30 Query: 61 QPSFGIAKKGHESRVIFSGPLMGHEFTSFILALLQVSGRAPKVEDEDIKRIKGIEKTINL 120
 PSF IAKKG ES V F+GLP+GHEFTSFILALLQVSGR PKV+ DI+KRI+ +++ ++
 Sbjct: 61 TPSFRIAKKGQESGVEFAGLPLGHEFTSFILALLQVSGRPPKVETDIVKRIQAVDEPMHF 120
- 35 Query: 121 ETYVSLTCHNCPDVVQAFNIMAVLNPNIHTMIEGGMQDEVKSKGIMSVPTVYKDQEEF 180
 ETYVSLTCHNCPDVVQAFNIM+V+NPNI+HTM+EGGM++DE+++KGIMSVPTVYKD EF
 Sbjct: 121 ETYVSLTCHNCPDVVQAFNIMSVVNPNIHTMVEGGMFKDEIEAKGIMSVPTVYKDGTEF 180
- 40 Query: 181 TSGRATIEQLLEQLDGLDAAFAFADKGVYDVLVIGGGPAGNSAAIYAARKGLKTGILAET 240
 TSGRA+IEQLL+ + GPL +AF DKGV+DVLVIGGGPAGNSAAIYAARKG+KTG+LAET
 Sbjct: 181 TSGRASIEQLLDLIAAGPLKEDAFDDKGVDFVLVIGGGPAGNSAAIYAARKGVKTGLLAET 240
- 45 Query: 241 FGGQVIETVGIENMIGTPLYTEGPKLMAQIEHTKSYDIDIKSQLATGIEKKELVEVTIA 300
 GGQV+ETVGIENMIGT Y EGP+LMAQ+EEHTKSY +DI+K+ A I+K +LVEV L
 Sbjct: 241 MGGQVMETVGIENMIGTPYVEGPQLMAQVEEHTKSYSDIMKAPRAKSIQRTDLVEVELD 300
- 50 Query: 301 NGAILQAKTAILALGAKWRNINVPGEFFRNKGVTYCPHCDGPLFEGKDVAVIGGGNSGM 360
 NGA L+AKTA+LALGAKWR INVPGE+EF NKGVTYCPHCDGPLF K VAVIGGGNSG+
 Sbjct: 301 NGAHLKAKTAVLALGAKWRKINVPGEKEFFNKGVTYCPHCDGPLFTDKKAVIGGGNSGL 360
- 55 Query: 361 EAAIDLAGVTKHVTVLEFLPELKADQVLQERAAKTDNLTKNVATKDIVGEDHVTGLNY 420
 EAA+DLAG+ HV +LEFLPELKAD++LQ+RA DN+TIL NVATK+I+G DHV GL Y
 Sbjct: 361 BAAIDLAGLASHVYILEFLPELKADKILQDRAEALDNITILT/NVATKEIIGNDHVEGLRY 420
- Query: 421 TDRDTNEEKHIDLEGVFVQIGLVPSTSWLKDSGIELNERQEIVVDKFGSTNIPGIFAAGD 480
 +DR TNEE +DLEGVFVQIGLVPST WLKDSG+ LNE+ EI+V K G+TNIP IFAAGD
 Sbjct: 421 SDRDTNEEYLLDLEGVFVQIGLVPSTDWLKDSGLALNEKGEIIVAKDGATNIPATFAAGD 480
- Query: 481 CTDAAYKQIIISMGS GATAAIGAFDYLR 509
 CTD+AYKQIIISMGS GATAA+GAFDYLR
 Sbjct: 481 CTD SAYKQIIISMGS GATAALGAFDYLR 509

-1622-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4513> which encodes the amino acid sequence <SEQ ID 4514>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0654(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 419/510 (82%), Positives = 472/510 (92%)

Query: 1 MVLDK E I K A Q L A Q Y L D L L E S D I V L Q A D L G D N D S Q K V K D F L D E I V A M S D R I S L E S T H L K R 60
M L + I K Q L A Q Y L L L E + D + V L Q L G D N + S Q K V K D F + + E I A M S + R I S + E + L R
Sbjct: 1 M A L S P D I K E Q L A Q Y L T L L E A D L V L Q V S L G D N E Q S Q K V K D F V E E I A A M S E R I S T E N I T L D R 60

Query: 61 Q P S F G I A K K G H E S R V I F S G L P M G H E F T S F I L A L L Q V S G R A P K V D E D I I K R I K G I E K T I N L 120
Q P S F + A K K G H S V + F + G L P + G H E T S F I L A L L Q V S G R A P K V D + D + I R I K I + + + +
Sbjct: 61 Q P S F K V A K K G H G S G V F A G L P L G H E L T S F I L A L L Q V S G R A P K V D Q D V I D R I K A I D R P L H F 120

Query: 121 E T Y V S L T C H N C P D V V Q A F N I M A V L N P N I T H T M I E G G M Y Q D E V K S K G I M S V P T V Y K D Q E E F 180
E T Y V S L T C H N C P D V V Q A N I M + V L N I + H T M + E G G M + Q D E V K + K G I M S V P T V + D E E F
Sbjct: 121 E T Y V S L T C H N C P D V V Q A L N I M S V L N D K I S H T M V E G G M F Q D E V K A K G I M S V P T V F L D G E E F 180

Query: 181 T S G R A T I E Q L L E Q L D G P L D A E A F A D K G V Y D V L V I G G G P A G N S A A I Y A A R K G L K T G I L A E T 240
T S G R A T I E Q L L E Q + G P L E A F A D K G + Y D V L V I G G G P A G N S A A I Y A A R K G L K T G + L A E T
Sbjct: 181 T S G R A T I E Q L L E Q I A G P L S E E A F A D K G L Y D V L V I G G G P A G N S A A I Y A A R K G L K T G L L A E T 240

Query: 241 F G G Q V I E T V G I E N M I G T L Y T E G P K L M A Q I E H T K S Y D I D I I K S Q L A T G I E K K E L V E V T L A 300
F G G Q V + E T V G I E N M I G T L Y T E G P K L M A + + E H T K S Y D + D I I K + Q L A T I E K K E + E V T L A
Sbjct: 241 F G G Q V M E T V G I E N M I G T L Y T E G P K L M A E V A H T K S Y D V D I I K A Q L A T S I E K K E N I E V T L A 300

Query: 301 N G A I L Q A K T A I L A L G A K W R N I N V P G E E F R N K G V T Y C P H C D G P L F E G K D V A V I G G G N S G M 360
N G A + L Q A K T A I L A L G A K W R N I N V P G E + E F R N K G V T Y C P H C D G P L F E G K D V A V I G G G N S G +
Sbjct: 301 N G A V L Q A K T A I L A L G A K W R N I N V P G E D E F R N K G V T Y C P H C D G P L F E G K D V A V I G G G N S G L 360

Query: 361 E A A L D L A G V T K H V T V L E F L P E L K A D Q V L Q E R A A K T D N L T I L K N V A T K D I V G E D H V T G L N Y 420
E A A L D L A G + K H V V L E F L P E L K A D + V L Q + R A A K T + N + T I + K N V A T K D I V G E D H V T G L N Y
Sbjct: 361 E A A L D L A G L A K H V Y V L E F L P E L K A D K V L Q D R A A K T N N M T I I K N V A T K D I V G E D H V T G L N Y 420

Query: 421 T D R D T N E E K H I D L E G V F V Q I G L V P S T S W L K D S G I E L N E R Q E I V V D K F G S T N I P G I F A A G D 480
T + R D + E + K H + D L E G V F V Q I G L V P + T + W L K D S G + L + R E I + V D K G S T N I P G I F A A G D
Sbjct: 421 T E R D S G E D K H L D L E G V F V Q I G L V P N T A W L K D S G V N L T D R G E I I V D K H G S T N I P G I F A A G D 480

Query: 481 C T D A A Y K Q I I I S M G S G A T A A I G A F D Y L I R Q 510
C T D + A Y K Q I I I S M G S G A T A A I G A F D Y L I R Q
Sbjct: 481 C T D S A Y K Q I I I S M G S G A T A A I G A F D Y L I R Q 510

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1470

A DNA sequence (GBSx1556) was identified in *S.agalactiae* <SEQ ID 4515> which encodes the amino acid sequence <SEQ ID 4516>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2906(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1623-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BAA25695 GB:AB010712 alkyl hydroperoxidase [Streptococcus mutans]
Identities = 167/186 (89%), Positives = 179/186 (95%)

Query: 1 MSLVGKEIIIEFSAQAYHDGKFITVTNEDVKGKWAVFCFYPADFSFVCPTLGDLEQYET 60
MSLVGKE++EFSAQAYH G+F+TV NEDVKGKWAVFCFYPADFSFVCPTLGDLEQY T
Sbjct: 1 MSLVGKEMVEFSAQAYHQGEFVTNNEDVKGKWAVFCFYPADFSFVCPTLGDLEQYAT 60

10 Query: 61 LKSLDVEVYSVSTDTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQGFVDLQDGLAQRG 120
L+SL VEVYSVSTDTHFVHKAWHDDSDVVGTTITY MIGDPSH++SQGF+VLG+DGLAQRG
Sbjct: 61 LQSLGVEVYSVSTDTHFVHKAWHDDSDVVGTTITYTMIGDPSHVLISQGFVVLGEDGLAQRG 120

15 Query: 121 TFIIDPDGVIQMMEINADGIGRDASTLIDKVRAAQYIRQHTGEVCPAKWKEGAETLTPSL 180
TFI+DPDG+IQMME+NADGIGRDASTLIDKVRAAQYIRQH GEVCPAKWKEGAETL PSL
Sbjct: 121 TFIVDPDGIIQMMEVNADGIGRDASTLIDKVRAAQYIRQHPGEVCPAKWKEGAETLKPSL 180

20 Query: 181 DLVGKI 186
DLVGKI
Sbjct: 181 DLVGKI 186

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4517> which encodes the amino acid sequence <SEQ ID 4518>. Analysis of this protein sequence reveals the following:

25 Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3022(Affirmative) < succ>
30 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 173/186 (93%), Positives = 181/186 (97%)

35 Query: 1 MSLVGKEIIIEFSAQAYHDGKFITVTNEDVKGKWAVFCFYPADFSFVCPTLGDLEQYET 60
MSL+GKEI EFSQAYHDGKFITVTNEDVKGKWAVFCFYPADFSFVCPTLGDLEQYET
Sbjct: 1 MSLIGKEIAEFSAQAYHDGKFITVTNEDVKGKWAVFCFYPADFSFVCPTLGDLEQYET 60

40 Query: 61 LKSLDVEVYSVSTDTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQGFVDLQDGLAQRG 120
LKSL VEVYSVSTDTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQ F+VLG+DGLAQRG
Sbjct: 61 LKSLGVEVYSVSTDTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQAFEVLGEDGLAQRG 120

45 Query: 121 TFIIDPDGVIQMMEINADGIGRDASTLIDKVRAAQYIRQHTGEVCPAKWKEGAETLTPSL 180
TFI+DPDG+IQMMEINADGIGRDASTLIDK+ AAQY+R+H GEVCPAKWKEGAETLTPSL
Sbjct: 121 TFIVDPDGIIQMMEINADGIGRDASTLIDKIHAQYVRKHGEVCPAKWKEGAETLTPSL 180

Query: 181 DLVGKI 186
DLVGKI
50 Sbjct: 181 DLVGKI 186

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1471

55 A DNA sequence (GBSx1557) was identified in *S.agalactiae* <SEQ ID 4519> which encodes the amino acid sequence <SEQ ID 4520>. This protein is predicted to be 30S ribosomal protein S2 (rpsB). Analysis of this protein sequence reveals the following:

Possible site: 60

-1624-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4462(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA50276 GB:X70925 30S ribosomal protein [Pediococcus
 acidilactici]
 Identities = 190/260 (73%), Positives = 226/260 (86%), Gaps = 4/260 (1%)

Query: 1 MAVISMKQLLEAGVHFGHQTRRWNPMAKYIFTERNGIHVIDLQQTIVKLADQAYEFVRDA 60
 M+VISMKQLLEAGVHFGHQTRRWNPKM +IFTERNGI++IDLQ+TVKL D AY FV+D

15 Sbjct: 1 MSVISMKQLLEAGVHFGHQTRRWNPKMKPFIFTERNGIYIIDLQKTIVKLIDNAYNFVKDV 60

Query: 61 AANDAVILFVGTKKQAAEAVAEAEAKRAGQYFINHRWLGGLTNWGTIQKRIARLKEIKRM 120
 AAND V+LFVGTKKQA A+ EEAKRAGQ+++NHRWLGGLTNW TIQKRI RLK++K+M

20 Sbjct: 61 AANDGVLEFVGTKKQAQTAIEEAKRAGQFYVNRWLGGLTNWNTIQKRIKRLKDLKRM 120

Query: 121 EEEGTFELLPKKEVALLNKQARLEKFLGGIEDMPRIPDVMYVVDPHKEQIAVKEAKKLG 180
 EE+GTF+ LPKKEVALLNKQ+ +LEKFLGGIEDMP IPDV++VVDP KEQIA+KEA+KL

25 Sbjct: 121 EEDGTFDRLPKKEVALLNKQKDKLEKFLGGIEDMPHIPDVLFVVDPRKEQIAIKEAQLN 180

Query: 181 IPVVMVDTNADPDDIDVIIIPANDDAIRAVKLITSKLADAVIEGRQGEDADV---DFAQ 236
 IPVVMVDTN DPD +DVIIP+NDDAIRAV+LITSK+ADAV+EGROGED + + A+

30 Sbjct: 181 IPVVMVDTNTDPDQVDVIIPSNDDAIRAVRLITSKMADAVVEGRQGEDDEAVQEEVAE 240

Query: 237 EAQADSIEEIVEVVEGSNND 256
 DS+E++ + VE +N+

30 Sbjct: 241 GVSKDSLEDLKTVEEGSNE 260

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4521> which encodes the amino acid
 sequence <SEQ ID 4522>. Analysis of this protein sequence reveals the following:

35 Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.4462(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 241/254 (94%), Positives = 248/254 (96%)

45 Query: 1 MAVISMKQLLEAGVHFGHQTRRWNPMAKYIFTERNGIHVIDLQQTIVKLADQAYEFVRDA 60
 MAVISMKQLLEAGVHFGHQTRRWNPMAKYIFTERNGIHVIDLQQTIVKLADQAYEFVRDA
 Sbjct: 1 MAVISMKQLLEAGVHFGHQTRRWNPMAKYIFTERNGIHVIDLQQTIVKLADQAYEFVRDA 60

50 Query: 61 AANDAVILFVGTKKQAAEAVAEAEAKRAGQYFINHRWLGGLTNWGTIQKRIARLKEIKRM 120
 AANDAVILFVGTKKQAAEAVA+EA RAGQYFINHRWLGGLTNWGTIQKRIARLKEIKRM
 Sbjct: 61 AANDAVILFVGTKKQAAEAVADEATRAGQYFINHRWLGGLTNWGTIQKRIARLKEIKRM 120

55 Query: 121 EEEGTFELLPKKEVALLNKQARLEKFLGGIEDMPRIPDVMYVVDPHKEQIAVKEAKKLG 180
 EEEGTF++LPKKEVALLNKQARLEKFLGGIEDMPRIPDVMYVVDPHKEQIAVKEAKKLG
 Sbjct: 121 EEEGTFDVLPEKKEVALLNKQARLEKFLGGIEDMPRIPDVMYVVDPHKEQIAVKEAKKLG 180

Query: 181 IPVVMVDTNADPDDIDVIIIPANDDAIRAVKLITSKLADAVIEGRQGEDADVDFACEAQA 240
 IPVVMVDTNADPDDID+IIPANDDAIRAVKLIT+KLADA+IEGRQGEDADV F + QA

60 Sbjct: 181 IPVVMVDTNADPDDIDIIIPANDDAIRAVKLITAKLADAIIEGRQGEDADVAFEADTQA 240

Query: 241 DSIEEIVEVVEGSN 254
 DSIEEIVEVVEG N

Sbjct: 241 DSIEEIVEVVEGDN 254

-1625-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1472

- 5 A DNA sequence (GBSx1558) was identified in *S.agalactiae* <SEQ ID 4523> which encodes the amino acid sequence <SEQ ID 4524>. Analysis of this protein sequence reveals the following:

Possible site: 49
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2648(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB73435 GB:AL139077 elongation factor TS [Campylobacter jejuni]
Identities = 169/358 (47%), Positives = 226/358 (62%), Gaps = 19/358 (5%)

20 Query: 1 MAEITAKLVKELREKSGAGVMDAKKALVETDGDLDKAIELLREKGMKA AAKKADRVA AEG 60
M EITA +VKELRE +GAG+MD K AL ET+GD DKA++LLREKG+ KAAKKADR+AAEG
Sbjct: 1 MTEITAAMVKELRESTGAGMMDCKNALSETNGDFDKAVQLLREKGLGKA AAKKADRLAAEG 60

25 Query: 61 LTGVYV--DGNVAAVIEVNAETDFVAKNDQFVTLVNETAKVIAEGRPSNNEEALALIMPS 118
L V V D A V E+N+ETDFVAKNDQF+ L +T I + EE + T+ +
Sbjct: 61 LVSVKVSDDFTSATVSEINSETDFVAKNDQFIALTKDITTAHIQSNLSQSV EELHSSTI-N 119

30 Query: 119 GETLEQAFVTATATIGEEKISFRRFALVEKTDEQHFQYQHNGGRIGVITV-----VEG 171
G E+ + ATIGE + RRFA ++ Y H GR+GV+ V
Sbjct: 120 GVKFEEYLLKSIATIGENLVRRFATLKAGANGVVNGYIHTNGRVGVVIAAACDSAEVAS 179

35 Query: 172 GDDALAKQVSMHVAAMKPTVLSYTELDAQFVHDELAQLNHHKIBQDNESRAMV---NKEAL 228
L +Q+ MH+AAM+P+ LSY +LD FV +E L ++E++NE R + NKP
Sbjct: 180 KSRDLLRQICMHIAAMRPSYLSYEDLDMTFVENEYKALVABELEKENEERRRLKDPNKP E H 239

40 Query: 229 PFLKYGSKAQLTDEVIAQA EEDIKAE LA AEGKPEKIWDKIVPGKMDRFLDNTKVDQEYT 288
++ S+ QL+D ++ +AEE IK EL A+GKPEKIWD I+PGKM+ F+ DN+++D + T
Sbjct: 240 KIPQFASRKQLSDAILKEAEEKIKEELKAQGGKPEKIWDNIIPGKMNSFIADNSQLDSKLT 299

45 Query: 289 LLAQVYIMDDSKTVEAYLESV-----NAKAVAFVRFEVGE GIEKASNDFAEVAATM 340
L+ Q Y+MDD KTVE + K V F+ FEVGE G+EK + DF AEVAA +
Sbjct: 300 LMGQFVVMDDKKTVEQVIAEKEKEFGGKIKIVFICFEVGE GLEKKTEDFAAEVAAQL 357

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4525> which encodes the amino acid sequence <SEQ ID 4526>. Analysis of this protein sequence reveals the following:

45 Possible site: 38
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3942(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

55 Identities = 307/344 (89%), Positives = 327/344 (94%)

Query: 1 MAEITAKLVKELREKSGAGVMDAKKALVETDGDLDKAIELLREKGMKA AAKKADRVA AEG 60
MAEITAKLVKELREKSGAGVMDAKKALVETDGD+DKA+ELLREKGMKA AAKKADRVA AEG
Sbjct: 33 MAEITAKLVKELREKSGAGVMDAKKALVETDGDMDKAVELLREKGMKA AAKKADRVA AEG 92

-1626-

Query: 61 LTGVYVDGNVAIVAEVNAETDFVAKNDQFVTLVNETAKVIAEGRPSNNBEALALTMPSGE 120
 LTGVYV GNVAIV+EVNAETDFVAKN QFV LVN TAKVIAEG+P+NN+EALAL MPSGE
 Sbjct: 93 LTGVYVHGNVAIVVEVNAETDFVAKNAQFVELVNATAKVIAEGKPANDEALALVMPSGE 152

5 Query: 121 TLEQAFVTATATIGEKISFRRFALVEKTDEQHFQAYQHNGGRIGVITVVEGGDDALAKQV 180
 TL +A+V ATATIGEKISFRRFAL+EK DEQHFQAYQHNGGRIGVI+VVEGGDDALAKQV
 Sbjct: 153 TLAEAYVNATATIGEKISFRRFALIEKADEQHFQAYQHNGGRIGVISVVEGGDDALAKQV 212

10 Query: 181 SMHVAAMKPTVLSYTELDAQFVHDELAQLNHNKIBQDNESRAMVNKFPALPFLKYGSKAQLT 240
 SMH+AAMKPTVLSYTELDAQF+ DELAQLNH IE DNESRAMV+KFPALPFLKYGSKAQL+
 Sbjct: 213 SMHIAAMKPTVLSYTELDAQFIKDELAQLNHAIELDNESRAMVDKFPALPFLKYGSKAQLS 272

15 Query: 241 DEVIAQAEEDIKAEALAEKGKPEKIWDKIVPGKMDRFMLDNTKVDQYETLLAQVYIMDDSK 300
 D+VI AE DIKAEALAEKGKPEKIWDKI+PGKMDRFMLDNTKVDQ YTLAQVYIMDDSK
 Sbjct: 273 DDVITAAEADIKAEALAEKGKPEKIWDKIIPGKMDRFMLDNTKVDQAYTLLAQVYIMDDSK 332

20 Query: 301 TVEAYLESVNAKAVAFVRFEVGEIEKASNDFEAEVAATMAAAL 344
 TVEAYL+SVNAKA+AF RFEVGEIEK +NDFE+EVAATMAAAL
 Sbjct: 333 TVEAYLDSVNAKAIAFARFEVGEIEKKANDFESEVAATMAAAL 376

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1473

A DNA sequence (GBSx1559) was identified in *S.agalactiae* <SEQ ID 4527> which encodes the amino acid sequence <SEQ ID 4528>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.1312(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1474

A DNA sequence (GBSx1560) was identified in *S.agalactiae* <SEQ ID 4529> which encodes the amino acid sequence <SEQ ID 4530>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have a cleavable N-term signal seq.

45 INTEGRAL Likelihood = -7.86 Transmembrane 128 - 144 (124 - 152)
 INTEGRAL Likelihood = -4.57 Transmembrane 35 - 51 (33 - 53)
 INTEGRAL Likelihood = -4.04 Transmembrane 92 - 108 (87 - 111)

----- Final Results -----

50 bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04953 GB:AP001511 small multidrug export related protein
 [Bacillus halodurans]

Identities = 47/137 (34%), Positives = 71/137 (51%), Gaps = 5/137 (3%)

+P+VELRG +P + G+ WEAL G++GN+LP+ I R + W + + +

Sbjct: 1 MPIVELRGGIPLGVVLGLSPWEALLFGIIGNLLPIVPILLLFRPISGWMLRFKQYQRLYD 60

W + +EK I L+LF +PLP TGA++ LAA L F+ + AV

Sbjct: 61 WLYNRTMKKSNNVEKFGA-----IGLILFTAVPLPTTGAYSACLA AVLFFIPFRFAFFAV 115

Query: 132 MLGVILAGCIMGTLII 148

GV++AG +M S I

Subject: 116 SAGVVIAGIVMTLFSYI 132

A related GBS gene <SEQ ID 8817> and protein <SEQ ID 8818> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0

McG: Discrim Score: 3.98

GvH: Signal Score (-7.5): -2.35

Possible site: 26

```
>>> Seems to have a cleavable N-term signal seq.
```

ALOM program count: 3 value: -7.86 threshold: 0.0

INTEGRAL Likelihood = -7.86 Transmembrane 128 - 144 (124 - 152)

INTEGRAL Likelihood = -4.57 Transmembrane 35 - 51 (33 - 53)

INTEGRAL	Likelihood = -4.04	Transmembrane	92 - 108 (87 - 111)
----------	--------------------	---------------	----------------------

PERIPHERAL	Likelihood = 12.20	109
------------	--------------------	-----

modified ALOM score: 2.07

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4142 (Affirmative) < succ>

```
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

LPXTG motif: 105-109

The protein has homology with the following sequences in the databases:

186 216 246 276 306 336 366 396
LTLIIISNF*KIRK*NLKSKTRMTADFESCHY*KOKIKWNTTIERFYLMNYIITFLISMIPLVELRGAVPFAIANGIPLM

: | : | | | | : | : | : |

GAIPALYFCFSPA

10 20

426 456 486 516 546 576 591 621
EALAIGVVGNNMLPVPPIIFFFARKVLEWGADKPYPYTGKFFTWCLKKGHSGGQKLEKVAGEKGL-----FIALLLVFGIPLPG

$$\begin{array}{ccccccc} | & : & | & | & | & | & \\ | & : & | & | & | & | & \end{array}$$

EAYLLSVLGNILPVPFLLFLDYLVRIATKVELLARIYR-----RVVERVERRKGVVERYGYLGLTIFVAIPLPV

40 50 60 70 80 90

651 681 711 741 771 801 831 861

TGAWTGTLAASLLDWEFKHSVIAVMLGVILAGCIMGTLSTIIGFNLF*KS*GEMTVSPF*YLPQHFDISKIRHLT*AKCLI

TGAWTGTLAFLQLNRLKAFLFISAGVCIAGVVLLASIGIIRLL

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1628-

Example 1475

A DNA sequence (GBSx1561) was identified in *S. galactiae* <SEQ ID 4531> which encodes the amino acid sequence <SEQ ID 4532>. This protein is predicted to be CtsR protein (ctsR). Analysis of this protein sequence reveals the following:

5 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.3672(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAB91548 GB:AJ249133 CtsR protein [Lactococcus lactis]
 Identities = 74/146 (50%), Positives = 103/146 (69%), Gaps = 3/146 (2%)

 Query: 4 KNTSDNIEEYIKSLLEQSGIAEIKRSNLADTFQVVPSQINIVIKTRFTESRGYVVESKRG 63
 KNTSD IE Y++ LLE++ + EIKR++LA+ F VVPSQINIVIKTRFT S+G+ VESKRG
 20 Sbjct: 5 KNTSDIIEAYLRQLEEAQVIEIKRADLANQFDVVPSQINIVIKTRFTASKGFDVESKRG 64

 Query: 64 GGGYIRIAKVHFSQHQHQLFGNMLSTIGERISEQVFDDLIQLLFDEEIIITEREGNLILATS 123
 GGGYI+I K +S +H+ + +S + D++QLLFDE+++TEREGNL+L
 25 Sbjct: 65 GGGYIKIVKYQSARHEFLTALYQKVPANLSSKAHDIQQLLFDEKVLITEREGNLLLLVI 124

 Query: 124 GDDVLGEQASVIRARMLRKLQLRDLR 149
 D G + R M++ ++ RLDR
 30 Sbjct: 125 TD---GAISPFTRGIMMKSIINRLDR 147

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4533> which encodes the amino acid sequence <SEQ ID 4534>. Analysis of this protein sequence reveals the following:

 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.2514(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

40 Identities = 117/151 (77%), Positives = 131/151 (86%)

 Query: 1 MAIKNTSDNIEEYIKSLLEQSGIAEIKRSNLADTFQVVPSQINIVIKTRFTESRGYVVES 60
 M KNTSD+IEEYIK LL +SGIAEIKRS LAD+QVVPSQINIVIKTRFTESRGY VES
 45 Sbjct: 1 MPTKNTSDSIEEYIKELLAKSGIAEIKRSM LADSFOVVPSQINIVIKTRFTESRGYEVE 60

 Query: 61 KRGGGGYIRIAKVHFSQHQHQLFGNMLSTIGERISEQVFDDLIQLLFDEEIIITEREGNLIL 120
 KRGGGGYIRIAKVHFSQ+H L GN+++TI + ISEQVF D IQLLFDE ++TEREGN+IL
 50 Sbjct: 61 KRGGGGYIRIAKVHFSQKHHLIGNLMTIEDCISEQVF TDSIQQLLFDEHLLTEREGNIIL 120

 Query: 121 ATSGDDVLGEQASVIRARMLRKLQLRDLRKG 151
 A + DDVLG S IRARML +LLQR+DRKG
 55 Sbjct: 121 AVASDDVLGTDGSTIRARMLYRLLQLRIDRKG 151

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1629-

Example 1476

A DNA sequence (GBSx1562) was identified in *S.agalactiae* <SEQ ID 4535> which encodes the amino acid sequence <SEQ ID 4536>. This protein is predicted to be ClpC (clpB-1). Analysis of this protein sequence reveals the following:

```

5   Possible site: 49
   >>> Seems to have no N-terminal signal sequence
       INTEGRAL    Likelihood = -2.34    Transmembrane    32 - 48 ( 32 - 49)

10  ----- Final Results -----
       bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
       bacterial outside --- Certainty=0.0000(Not Clear) < succ>
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAD01783 GB:AF023422 ClpC [Lactococcus lactis]
    Identities = 401/831 (48%), Positives = 571/831 (68%), Gaps = 52/831 (6%)

    Query: 4   YSIKIQEVFRLAQFQAARYESHYLESWHLLLAMVLVHDSVAGLTFAYE---SEVAIEEY 60
              Y+ L +F A A +Y+ +ES HLL AM S+A A S++ I+
20  Sbjet: 8   YTPTLDRIFEKAAEYAHQYQYGTIESAHLAAMATTSGSIAYSILAGMNVDSDDLIDLE 67

    Query: 61  EAATILALGRAPKEEITNYQFLEQSPALKKILKLAENISIVVGAEDVGTEHVLLAMLVNK 120
              + ++ + + R+ L SP ++++ +A +++ AE VGTEH+L A+L +
25  Sbjet: 68  DLSSHVKVKRSE-----LRFSPEAEVVTVASFLAVHNNAAEAVGTEHLLYALLQVE 118

    Query: 121 DLLATRILELVGFRGQDDGESVRMVDLRKALERHAGF-TKDDIKAIYELRNPKKAKSGAS 179
              D ++L+L + + +V LRK +E+ G ++ KA+ + K AK A
25  Sbjet: 119 DGFGLQLLKL-----QKINIVSLRKEIEKRTGLIVPENKKAVTPMSKRKMAKGVAE 169

    Query: 180 FSDMMKPPSTAGDLADFTDRDLQMAVDGEIEPVIGRDKEISRMVQVLSRKTNNPVLVGD 239
              S+ L + DL++ A G+++P+IGR+ E+ R++ +LSR+TKNNPVLVG+
30  Sbjet: 170 -----NSSTPTLDSVSSDLTEAARSGKLDPMIGREAEVDRLIHILSRRTKNNPVLVGE 222

    Query: 240 AGVGKTALAYGLAQRIANGNIPYELRDMRVLELDMSVAVGTRFRGDFEERMNQIADIE 299
              GVVGK+A+ GLAQRI NG +P L + R++ L+M +VVAGT+FRG+FE+R+ I+ ++
35  Sbjet: 223 PGVGKSAIIEGLAQRIVNGQVPIGLMNSRIMALNMATVVAGTKFRGEFEDRLTAIVEEVS 282

    Query: 300 EDGHILFIDELHTIMSGSGIDSTLDAANILKPALARGTLRTVGATTQEEYQKHIEKDA 359
              D ++FIDELHTI+G+G G+DS DAANILKPALARG + VGATT EYQK+IEKD
40  Sbjet: 283 ADPDVIIFIDELHTIIGAGGGMDSVNDAAANILKPALARGDFQMVGATTYHEYQKYIEKDE 342

    Query: 360 ALSRRFAKVLVEEPNLEDAVEILLGLKPAYEAFHNVTISDEAVMTAVKVAHRYLTSKNLP 419
              AL RR A++ V+EP+ ++A IL GL+ +E +H V +D+A+ +AV ++ RY+TS+ LP
45  Sbjet: 343 ALERRLARINVDSPDEAIAILLQGLREKFEDYHGVKFTDQAISAVTLSVRYMTSRKLP 402

    Query: 420 DSAIDLLEASATVQMMIKKNAPSLT-----EVDQAILDDDMKSA----- 460
              D AIDLLEA+A V++++K ++ E+ +A++ D+K++
45  Sbjet: 403 DKAILDLDEAAARVKILLKTKQNVFELEKDFVKAQEELAEAVIKLDVKASRIKEKAVEK 462

    Query: 461 --SKALKASYKGKKRKPIAVTEDHIMATLSRLSGIPVEKLTQADSKKYLNLEKELHKRVI 518
              K K S K +KR+ VT+ ++A S L+G+P+ ++T+++S + +NLEKELHKRV+
50  Sbjet: 463 ISDKIYKFSIKEERQE--VTDQAVIAVASTLTGVPITQMTKSESDRLINLEKELHKRVV 520

    Query: 519 GQDDAVTAISRARRNQSGIRTGKRPIGSMFLGPTGVGKTELAKALAEVLFDDDESALIR 578
              GQ++A++A+SRAIRR +SG+ +RP+GSFMFLGPTGVGKTELAKALA+ +F E +IR
55  Sbjet: 521 GQEEAISAVSRAIRARSGVADSRPPMGSMFLGPTGVGKTELAKALADSVFGSEDNMIR 580

    Query: 579 FDMSEYMEKFAASHLNGAPPGYVGYDEGGELTEKVRNKPYSVLLFDEVEKAHPDIFNVLL 638
              DMSE+MEK + S L GAPPGYVGYDEGG+LFE+VRNKPYSV+L DEVEKAH D+FN++L
60  Sbjet: 581 VDMSEFMEKHSTSRIGAPPGYVGYDEGGQLTERVRNKPYSVLLDEVEKAHLDVFNIML 640

    Query: 639 QVLDDGVLTDGRGRKVDFTNTIIIMTSNLGATALRDDKTGVFGAKDISHDYTAMQKRIME 698
              Q+LDDG +TD++GRKVDFTNTIIIMTSNLGATALRDDKTGVFGAK+I+ DY+AMQ RI+E
60  Sbjet: 641 QILDDGFVTDTKGRKVDFTNTIIIMTSNLGATALRDDKTGVFGAKNITADYSAMQSRILE 700

```

-1630-

Query: 699 ELKKAYRPEFINRIDEKVVFHSLSDNMREVVKIMVKPLILALKDKGMDLKFQPSALKHL 758
 ELK+ YRPEF+NRIDE +VFHSL + ++VKIM K LI L ++ + +K PSA+K +
 Sbjct: 701 ELKRHYRPEFLNRIDENIVFHSLSQEIEQIVKIMSKSLIKRLAEQDIHVKLTPSAIKLI 760

Query: 759 AEDGYDIEMGARPLRRTIQTQVEDHLSELLANQVKEGQVIKIGVSKGKLK 809
 AE G+D E GARPLR+ +Q +VED LSE LL+ ++K G I IG S K+K
 Sbjct: 761 AEVGFDPEYGARPLRKALQKEVEDLLSEQLLSGEIKAGNHISIGASNKKIK 811

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4537> which encodes the amino acid sequence <SEQ ID 4538>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence
 15 INTEGRAL Likelihood = -1.75 Transmembrane 32 - 48 (32 - 48)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 285-287

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 618/814 (75%), Positives = 716/814 (87%), Gaps = 1/814 (0%)

Query: 1 MSHYSIKLQEVFRLAQFQAARYESHYLESWHLLAMVLVHDSVAGLTFAYEYSEVAIEEY 60
 M YS K+Q++FR AQFQAAR++SH LE+WH+LLAMV V +S+A + +EY+++VAIEEY
 30 Sbjct: 1 MIMYSTKMQDIFRQAQFQAARFDSHCLETWHVLLAMVAVDNSLANMILSEYDAQVAIEEY 60

Query: 61 EAATILALGRAPKEEITNYQFLEQSPALKKILKLAENISIVVGAEDVGTEHVLLAMLVKN 120
 EAA IIA+G+ PKE+++ F QS L +L A+ IS + ++VG+EHVL A+L+N
 35 Sbjct: 61 EAAAILAMGKTPKEQLSRVDFRPQSKTLTNLLAFAQAISQITRDQEVGSEHVLFAILLNP 120

Query: 121 DLLATRILELVGFRGQDDGESV-RMVDLRKALERHAGFTKDDIKAIYELRNPKKAKSGAS 179
 D++A+R+LE+ G++ +D+G R+ DLRKA+ERHAG++K+ IKAI+ELR PKK K+ +
 40 Sbjct: 121 DIMASRLLEIAGYQIKDNGNQPRDLADLRKAIERHAGYSKEMIKAIHELKPKKTKTQGT 180

Query: 180 FSDMMKPPSTAGDLADFTRDLSQMAVDGEIEFPVIGRDKEISRMVQVLSRKTKNNPVLVGD 239
 FSDMMKPPSTAG+L+DFTRDL+MA G +E VIGRD+E+SRM+QVLSRKTKNNPVLVGD
 45 Sbjct: 181 FSDMMKPPSTAGELSDFTDRDLTEMARQGLLESVIGRDQEVSRMIQVLSRKTKNNPVLVGD 240

Query: 240 AGVGKTALAYGLAQRIANGNIPYELRDMRVLELDMMSVVAGTRFRGDFEERMNQIIDIE 299
 AGVGKTALAYGLAQRIANG IPYEL++MRVLELDMMSVVAGTRFRGDFEERMNQIIDIE
 50 Sbjct: 241 AGVGKTALAYGLAQRIANGAIPYELKEMRVLELDMMSVVAGTRFRGDFEERMNQIIDIE 300

Query: 300 EDGHIILFIDELHTIMGSGSGIDSTLDAANILKPALARGTLRTVGATTQEYQKHIEKDA 359
 DG IILF+DELHTIMGSGSGIDSTLDAANILKPAL+RGTL VGATTQEYQKHIEKDA
 55 Sbjct: 301 ADGQIILFVDELHTIMGSGSGIDSTLDAANILKPALSRGTLHVMGATTQEYQKHIEKDA 360

Query: 360 ALSRRFAKVLVEEPNLEDAYEILLGLKPAYEAFHNVTISDEAVMTAVKVAHRYLTSKNLP 419
 ALSRRFAK+L+EEPN EDAY+IL+GLK +YE +HNV+IS+EAV TAVK+AHRYLTSKNLP
 60 Sbjct: 361 ALSRRFAKILIEEPNTEDAYQILMGLKLSYETYHNVSISNEAVKTAVKMAHRYLTSKNLP 420

Query: 420 DSAIDLLDEASATVQMMIKKNAPSLLTEVDQAILDDDMKSASKALKASYKGGKRKPIAVT 479
 DSAIDLLDEASA VQ M+KK+AP LT +DQA+++ DMK S+ L KG+ RKP VT
 65 Sbjct: 421 DSAIDLLDEASAAVQNMVKSAPETLTPIDQALINGDMKKVSRLLAKEAKGQMRKPTPV 480

Query: 480 EDHIMATLSRLSGIPVEKLTQADSKKYLNLKELHKKRVIGQDDAVTAISRIRRNQSGIR 539
 ED I+ATLS+LSGIP+EKLTQADSKKYLNLKELHKKRVIGQD AVTAISRIRRNQSGIR
 70 Sbjct: 481 EDDILATLSKLSGIPLEKLTQADSKKYLNLKELHKKRVIGQDAVTAISRIRRNQSGIR 540

Query: 540 TGRPIGFSFMFLGPTGVGKTELAKALAEVLFDDDESALIRFDMSEYMEKFAASHINGAPP 599
 TGRPIGFSFMFLGPTGVGKTELAKALAEVLFDD+ALIRFDMSEYMEKFAAS INGAPP
 75 Sbjct: 541 TGRPIGFSFMFLGPTGVGKTELAKALAEVLFDD+ALIRFDMSEYMEKFAASRLNGAPP 600

Query: 600 YVGYDEGGELTEKVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDDGVLTD SRGRKVD FSNT 659
YVGYDEGGELT+KVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDDG+LTD SRGRKVD FSNT
Sbjct: 601 YVGYDEGGELTQKVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDDGILTD SRGRKVD FSNT 660

Query: 660 IIIMTSNLGATALRDDKTVGFGAKDISHDYTA MQKRIMEELKKAYRPEFINRIDEKVVFH 719
IIIMTSNLGATALRDDKTVGFG KDI D+ AM+KRI+EEL+K YRPEFINRIDEKVVFH
Sbjct: 661 IIIMTSNLGATALRDDKTVGFGVKDIHQDHQAMEKRILEELRKTYRPEFINRIDEKVVFH 720

Query: 720 SLSDQDNMREVVKIMVKPLILALKDKGMDLKFPQSALKHLAEDGYDIEMGARPLRRTIQTO 779
SL+QDNMR+VVKIMV+PLI L +KG+ LK QP ALKHL+E GYD MGARPLRRT+QT+
Sbjct: 721 SLTQDNMRDVVKIMVQPLITTLAEKGITLKIQLPLALKHLSEVGYDEHMGARPLRRTLQTE 780

Query: 780 VEDHLSLELLLANQVKEGQVIKIGVSKGKLFIDIA 813
+ED LSEL+L+ ++ G +KIG+S GKL F IA
Sbjct: 781 IEDKLSLELILSRELTSGHTLKGILSHGKLT FPIIA 814

```

20      Lipop: Possible site: -1      Crend: 9
      McG: Discrim Score:      -13.52
      GvH: Signal Score (-7.5): -2.1
          Possible site: 49
      >>> Seems to have no N-terminal signal sequence
25      ALOM program      count: 1 value:      -2.34 threshold:      0.0
          INTEGRAL      Likelihood = -2.34      Transmembrane      32 -      48 (      32 -      49)
          PERIPHERAL      Likelihood =      0.95      112
      modified ALOM score:      0.97

30      *** Reasoning Step: 3

      ----- Final Results -----
          bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

47.4/69.6% over 804aa

Listeria monocytogenes

```

40      EGAD|136761| ClpC ATPase Insert characterized
       GP|1314297|gb|AAC44446.1||U40604 ClpC ATPase Insert characterized

ORF00207(298 - 2727 of 3045)
45    EGAD|136761|145854(2   -   806   of   825)   ClpC   ATPase   {Listeria monocytogenes}
       GP|1314297|gb|AAC44446.1||U40604 ClpC ATPase {Listeria monocytogenes}
       %Match = 33.6
       %Identity = 47.4 %Similarity = 69.6
       Matches = 372 Mismatches = 229 Conservative Sub.s = 174

50    87          117          147          177          207          237          267          297
       SFF*STPIIWKYVINDWRAYQ*TSF**FDSIIIR*RDNYRT*RKFDSGDIR**RLRRASLCY*SSYAP*IITITR*KRIP

                                                                    M

55    327          357          387          417          447          477          507          537
       FMSHYSIKLQEVFLRAQFQAARYESHYLESWHLLLAMVLVDHVSAGLTFAEYESSEVAIEEVEAATILALGRAPKEEITNY
           :: |:: |:: || | : : || : | : : | || :: |: : : | : : : |
       MFGRFTQRAQKVLALESQEEMARLNHSNLGTEHILLGLVREGEGIAA--KALVELGISSEKVKQOEVEGLIGHG-EKAVITTI
                   20            30            40            50            60            70

60    567          597          627          657          687          717          744          774
       QFLEQSPALKKITLKLAENISIVVGAEVDGTEHVLLAMLVNKDLLATRILXLVGFGRQDDGESV-RMVDLRKALERHAGFT
       |: | ||::|: : : : | ||||:| : : : | : | : | : | : : |
       QYT---PRAKKVIELSMDEARKLGHTYVGTEHHILLGLIREGEVGAARVLSNLTGISLNKARQQVLQLLGGGDA-----

65    90            100            110            120            130            140

```

[illegible]

SEQ ID 8820 (GBS26) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 9; MW 93.3kDa), in Figure 167 (lane 16 & 17; MW 108kDa) and in Figure 239 (lane 14; MW 108kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 7; MW 118kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1633-

Example 1477

A DNA sequence (GBSx1563) was identified in *S.galactiae* <SEQ ID 4539> which encodes the amino acid sequence <SEQ ID 4540>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4541> which encodes the amino acid sequence <SEQ ID 4542>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 178/213 (83%), Positives = 199/213 (92%)

Query: 1 MLIVLAGTIGAGKSSLAALGQHGLGTDVFYEAVDNNPVLDDYYQDPQKYAFLLQIIFLNK 60
 MLIVLAGTIGAGKSSLAALG+HLGTDVFYEAVDNNPVLDDYYQDP+KYAFLLQI+FLNK
 Sbjct: 1 MLIVLAGTIGAGKSSLAALGHEHLGTDVFYEAVDNNPVLDDYYQDPKKYAFLLQIYFLNK 60

Query: 61 RFQSIKEAYKANNVLDRSIFEDELFLTLNYKNGNVTKTELDIYKELLANMLEELEGMPK 120
 RF+SIKEAY+A+NN+LDRSIFEDELFL LNYKNGNVTKTELDIY+ELLANMLEELEGMPK
 Sbjct: 61 RFKSIKEAYQADNNILDRSIFEDELFLKLNKNGNVTKTELDIYQELLANMLEELEGMPK 120

Query: 121 KRPDLLVYIDVSFDKMLERIDKGRSFEQVDSNPEDYDYKQVHSEYPEWYENYDVSPKI 180
 KRPDLL+YIDVSFDKMLERI++RGRSFEQVD NP L YY QVH EYP WYE+Y+VSPK+
 Sbjct: 121 KRPDLLIYIDVSFDKMLERIERRGRSFEQVDGNPSLEQYYHGVHGEYPTWYEDYEVSPKM 180

Query: 181 RIDGNKLDVFNKPEDLQHVLDITIDSELQKLDLL 213
 +IDGN LDFV+NP+DL VL ID++L++L LL
 Sbjct: 181 KIDGNSLDFVQNPQDLATVLKMDITKLKELHLL 213

A related GBS gene <SEQ ID 8821> and protein <SEQ ID 8822> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0
 McG: Discrim Score: 3.94
 GvH: Signal Score (-7.5): 1.42
 Possible site: 17
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 7.69 threshold: 0.0
 PERIPHERAL Likelihood = 7.69 49
 modified ALOM score: -2.04

*** Reasoning Step: 3

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1634-

SEQ ID 4540 (GBS9) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 5; MW 52kDa) and Figure 12 (lane 2 & 3; MW 50.3kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 6; MW 27kDa) and Figure 3 (lane 2; MW 25kDa). The GBS9-GST fusion product was purified (Figure 191, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 318), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1478

10 A DNA sequence (GBSx1564) was identified in *S.agalactiae* <SEQ ID 4543> which encodes the amino acid sequence <SEQ ID 4544>. Analysis of this protein sequence reveals the following:

Possible site: 36
>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1182(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4545> which encodes the amino acid sequence <SEQ ID 4546>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 281/323 (86%), Positives = 305/323 (93%)

Query: 3 QLNSSFMIGKVEIPHRTVLAPMAGITNSAFRTIAKEFGAGLVVMMISEKGLLYNNEKTL 62
+LNSSF IG VEIPHRTVLAPMAG+TNSAFRTIAKEFGAGLVVMMISEKGLLYNNEKTL
35 Sbjct: 27 KLNSSFRIGDVEIPHRTVLAPMAGVTNSAFRTIAKEFGAGLVVMMISEKGLLYNNEKTL 86

Query: 63 HMLHIDENEHPMSIQLFGGDAEGLKRAADFIQSNTKADIVDINMGCPVNVKVKNEAGAKW 122
HMLHIDENEHPMSIQLFGGDAEGLKRAADFIQ+NTKADIVDINMGCPVNVKVKNEAGAKW
Sbjct: 87 HMLHIDENEHPMSIQLFGGDAEGLKRAADFIQTNTKADIVDINMGCPVNVKVKNEAGAKW 146

40 Query: 123 LRDPKIIYHIVKEVTSVLDIPLTVKMRTIGWSDSSNAIENALAAESAGVSALAMHGRTREQ 182
LRDP+KIIYHIVKEVTSVLDIPLTVKMRTIGW+DSS A+ENALAAESAGVSALAMHGRTREQ
Sbjct: 147 LRDPKIIYHIVKEVTSVLDIPLTVKMRTIGWADSSLAENALAAESAGVSALAMHGRTREQ 206

45 Query: 183 MYTGTCDHETLGKAVAVTSPFIANGDIRTVHDAKFMIIEIGADAIMVGRGARSNPYIF 242
MYTGTCDHETL +V+KA+T IPFI NCD+R+V DAKFMIEEIG DA+M+GR A +NPY+F
Sbjct: 207 MYTGTCDHETLARVSKAITKIPFIGNDVRSVQDAKFMIIEIGVDAVMIGRAAMNPNYLF 266

50 Query: 243 TQINHEFFETGEILPDLPEFKMLDVAEDHLTRLVNLKGETIAVREFRGLAPHYLRGKSGAA 302
TQINHEFFETG+ LPDLPF K LD+A+DHL RL+NLKGETIAVREFRGLAPHYLRG +GAA
Sbjct: 267 TQINHEFFETGQELPDLPEFAKKLDIAKDHLKRLINLKGETIAVREFRGLAPHYLRGTAGAA 326

Query: 303 KIRGAVSRAETLAEVQELFAGLR 325
K+RGAVSRAETLAEV+ +F +R
55 Sbjct: 327 KVRGAVSRAETLAEVEAIFETVR 349

-1635-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1479

A DNA sequence (GBSx1565) was identified in *S.agalactiae* <SEQ ID 4547> which encodes the amino acid sequence <SEQ ID 4548>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2164(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 3930:

15 Identities = 235/288 (81%), Positives = 259/288 (89%)

Query: 1 MDKIIKSISTSGSFRAYVLDCTETVTRTAQEKHQTLSSTVALGRTLIAHQILAANQKGNS 60
 MDKIIKSI+ SG+FRAYVLD TETV AQEKH TLSSTVALGRTLIAHQILAANQKG+S
Sbjct: 1 MDKIIKSIAQSGAFRAYVLDSTETVALAQEKHNTLSSTVALGRTLIAHQILAANQKGDS 60

20 Query: 61 KVTVKVIGDSSFGHIISVADTKGNVKGVIQNTGVDIKKTATGEVLVGPFGNGHFVVTID 120
 K+TVKVGIDSSFGHIISVADTKG+VKGVIQNTGVDIKKTATGEVLVGPFGNGHFV I D
Sbjct: 61 KITVKVIGDSSFGHIISVADTKGHVKGVIQNTGVDIKKTATGEVLVGPFGNGHFVVTIID 120

25 Query: 121 YATGQPYTSTTPLITGEIGEDFAYYLTESEQTPSAVGLNVLLDDEKVKVAGGFMLQVLP 180
 Y TG PYTSTTPLITGEIGEDFAYYLTESEQTPSA+GLNVLLD+ DKVKVAGGFMLQVLP
Sbjct: 121 YGTGNPYTSTTPLITGEIGEDFAYYLTESEQTPSAIGLVLLDENDKVKVAGGFMLQVLP 180

30 Query: 181 GASDEEISRYEKRIQEMPSISSLLESENHIESLLSAIYGEDDYKRLSEDSLAFYCDCSKE 240
 GAS+EEI+RYEKR+QEMP+IS LL S+NH+++LL AIYG++ YKRLSE+ L+F CDCS+E
Sbjct: 181 GASEEEIARYEKRIQEMPAISHLASKNHVDALLEAIYGDEPYKRLSEEPSLFCDCSRE 240

35 Query: 241 RFEAALLTLTKELQAMKDEKGVETCQFCNQTYTFTEEDLEKIIND 288
 RFEAAL+TL +LQAM DEDKG EI CQFC Y F E DLE II+D
Sbjct: 241 RFEAALMTLPKADLQAMIDEDKGABIVCQFCGTYQFNESDLEAIISD 288

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1480

40 A DNA sequence (GBSx1566) was identified in *S.agalactiae* <SEQ ID 4549> which encodes the amino acid sequence <SEQ ID 4550>. This protein is predicted to be surface-located membrane protein 1 (Imp1). Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.4312(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB93480 GB:AF019377 tellurite resistance protein [Rhodobacter
 sphaeroides]

55 Identities = 64/350 (18%), Positives = 146/350 (41%), Gaps = 7/350 (2%)

-1636-

Query: 44 LTPAQKSAISEKTPALVDTFVGDQNALDFGQSAVEGVNTTVNHILSEQKKIQIPQVDDL 103
 L A E + + V D + + + F G A + T + L + + K + D
 Sbjct: 34 LASAPPEKAQEIRRRMAELNVSDSQSIIGFGSKAQAEIQTISQQLADVKNKDVGPAGDS 93

5 Query: 104 LKNANRELNGFIAKYKDATPAELEKKPNLIQKLFKQSKTSLQEFYFDSQNIQKMDMMMA 163
 L+ + GF + + + K + + L + + F + + + + Q + + D +
 Sbjct: 94 LREVVSTIRGF-----SVSEFDVRRKASWVERLLGRT-APFARFVARYEDVQQQIDRITQ 147

10 Query: 164 NVVKQEDTLARNIVSAEMLIEDNTKSIENLVGVIAFIESSQAEANRASHLQOEILALDS 223
 + + + E L + + I + + L + L I A + A + R + + + A
 Sbjct: 148 SLLTHEHRLKDIKGLDILYARTLDFYDELALYIAAGDEVLDLGRVIPAKEAEVAATP 207

15 Query: 224 QTSEYQIKSNQLARMTEVINTLEQQHPEYVSRLYVAVATTPQMRNLVKVSSDMRQKLGML 283
 + + I K + + L + + L E + + + V + P + R + + + + +
 Sbjct: 208 E-GDRMIKAQELRDLRAARDLERRVHDLKLTRQVTMQSLPSIRLVQENDKALVTRINST 266

20 Query: 284 RRNTIPTMKLSIAQLGMMQQSVKSGVTADAIVNANNAALQMLAETSKEAIPMLEKTAQSP 343
 NT+P + +AQ +Q+S + + + N L AE + + A + + K +
 Sbjct: 267 LVNTVPLWETQLAQAVTQRSREAAEAVRGASDLTNEILITANAENLQQANKIVRKEMERG 326

Query: 344 TVSIKSVTALAESLVAQNNGIIAAIDKGRKERAQLES AVIKSAETINDSV 393
 I + + V + L + A N + A D + GR RA E + + + + D + +
 Sbjct: 327 VFDIEAVKKANATLIATINESLAIADegrARRATAETELQRMELRDL 376

- 25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4551> which encodes the amino acid sequence <SEQ ID 4552>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3230(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 35 An alignment of the GAS and GBS proteins is shown below.

Identities = 333/413 (80%), Positives = 379/413 (91%)

Query: 5 FNFIDIDQIADNAITKTDKTEIISNQTTSTGQIAFFEKLTPAQKSAISEKTPALVDTFV 64
 FNFIDIDQIADNA+ KTDKTT+IIS+ T GQI+FFEKL+ Q++AI+ K PALVDTF+
 40 Sbjct: 4 FNFIDIDQIADNAVIKTDKTTDIISDLPTDTNGQISFFEKLSDAQQTAITAKAPALVDTFL 63

Query: 65 GDQNALDFGQSAVEGVNTTVNHILSEQKKIQIPQVDDLKKNANRELNGFIAKYKDATPA 124
 DQNALDFGQSAVEGVN TVNHIL+EQKK+QIPQVDDLK+ NRELNGFIAKYKDATP
 45 Sbjct: 64 ADQNALDFGQSAVEGVNATVNHILAEQKKLQIPQVDDLKSTNRELNGFIAKYKDATPV 123

Query: 125 ELEKKPNLIQKLFKQSKTSLQEFYFDSQNIQKMDMMMAANVVKQEDTLARNIVSAEMLIE 184
 +L+KPN +QKLFKQ+ +LQEFYFDSQNIQKMD MAA VVKQEDTLARNIVSAE+LIE
 Sbjct: 124 DLDDKPNFLQKLFKQSRDTLQEFYFDSQNIQKMDMAAAVVKQEDTLARNIVSAELLIE 183

50 Query: 185 DNTKSIENLVGVIAFIESSQAEANRASHLQOEILALDSQTSEYQIKSNQLARMTEVINT 244
 DNTKSIENLVGVIAFIE+SQ EA+ RA+ LQ+++ DS T +YQIK++ LAR TEVINT
 Sbjct: 184 DNTKSIENLVGVIAFIEASQKEASQRAAALQKDLKTKDSATPDYQIKADLLARTTEVINT 243

55 Query: 245 LEQQHPEYVSRLYVAVATTPQMRNLVKVSSDMRQKLGMLRRNTIPTMKLSIAQLGMMQQS 304
 LEQQH EY+SRLYVAVATTPQMRNLVKVSSDMRQKLGMLRRNTIPTMKLSIAQLGMMQQS
 Sbjct: 244 LEQQHTEYLSRLYVAVATTPQMRNLVKVSSDMRQKLGMLRRNTIPTMKLSIAQLGMMQQS 303

Query: 305 VKSGVTADAIVNANNAALQMLAETSKEAIPMLEKTAQSP TVSIKSVTALAESLVAQNNGI 364
 VKSG+TADAI+NANNAALQMLAETSKEAIP LE++AQ+PT+S+KSVT+LAESLVAQNNGI
 60 Sbjct: 304 VKSGMTADAIINANNAALQMLAETSKEAIPALEQSAQNPTLSMKSVTSLAESLVAQNNGI 363

Query: 365 IAAIDKGRKERAQLES AVIKSAETINDSVKIRDKKIVEALLNEGKSTQEKVDE 417
 IAAID GRKERAQLESA+I+SAETINDSVK+RD+ IV+ALL+EGK TQ+ +D+
 65 Sbjct: 364 IAAIDHGRKERAQLESAIIRSAETINDSVKLRDQNIQVQALLSEKGTQKTIDK 416

-1637-

SEQ ID 4550 (GBS201) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 5; MW 49kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 3; MW 74.5kDa) and in Figure 62 (lane 8 & 9; MW 74.5kDa). The GBS201-GST fusion product was purified (Figure 209, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 304), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1481

A DNA sequence (GBSx1567) was identified in *S.galactiae* <SEQ ID 4553> which encodes the amino acid sequence <SEQ ID 4554>. This protein is predicted to be rhopty protein. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have an uncleavable N-term signal seq
15  INTEGRAL    Likelihood = -6.58    Transmembrane 13 - 29 ( 10 - 31)
    INTEGRAL    Likelihood = -1.54    Transmembrane 33 - 49 ( 33 - 49)

----- Final Results -----
20      bacterial membrane --- Certainty=0.3633(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4555> which encodes the amino acid sequence <SEQ ID 4556>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
30 ----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 115/239 (48%), Positives = 162/239 (67%), Gaps = 3/239 (1%)

40 Query: 32 EVIATLLIIGGGYCAYYVD-KKRLKRFTSNQRTEALKSDIKETDQDIRHLEILKKNRS 90
    +++ + I G GY + V +KRL + +++E LK+ I+ D+ +R L+ D+
Sbjct: 42 DILPAIAIGGTGYAIFRVRSHQKRLAKAKIAKQLEDLAKIQLADRKVRLDLYLADHDD 101

Query: 91 KEYIKLAHQILPQLDLIRNEANQLQKAIEPNYKRIKTKANTFSNEINEQLIKLHASP 150
    +Y LA Q+LPQL I+ +A L+ ++P IY+RITKAN ++I QL L + L
45 Sbjct: 102 FQYNVLAQQLLPQLSDIKAKAITLKDQLDPQIYRITKANDVESDITLQLETQIATTL 161

Query: 151 --EPISDQDEDEMIRIAPELKPFYHNIQDDHFAILKKIEADNKAELAAIHQANMKRFTDV 208
    +P+ +I APELKP+Y NIQ DH AIL KI+ ADN+ EL A+H ANM+RF D+
50 Sbjct: 162 NQQLKTPSPNLINKAPELKPYDNIQTQDHQAILAKIQGADNQEELLALHDANMRRFEDI 221

Query: 209 LAGYIRIKQSPKFNNAKERLEQALQAIKKFNLDLDETLRQLNESDMKDFDVSRLMMQ 267
    L GY++IK+ PKN+ NA RLEQA QAI++F+ DLDETLR+LINESD+KDFD+SLR+MQG
Sbjct: 222 LTGYLKIKEEPKNYNAARLEQAKQAIQQFDEDLDETLRRLNESDLKDFDISLRIMQ 280

```

-1638-

SEQ ID 4554 (GBS265) was expressed in *E. coli* as GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 2; MW 56kDa) and in Figure 62 (lane 6; MW 56.3kDa).

5 The GBS265-GST fusion product was purified (Figure 207, lane 5) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 258A) and FACS (Figure 258B). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1482

10 A DNA sequence (GBSx1568) was identified in *S. agalactiae* <SEQ ID 4557> which encodes the amino acid sequence <SEQ ID 4558>. This protein is predicted to be glutamate--cysteine ligase (gshA). Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence
 15 INTEGRAL Likelihood = -1.70 Transmembrane 575 - 591 (575 - 591)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG08588 GB:AE004933 glutamate--cysteine ligase [*Pseudomonas aeruginosa*]
 Identities = 142/468 (30%), Positives = 220/468 (46%), Gaps = 62/468 (13%)
 25 Query: 12 SHLPIL-QATFGLERESLRHQPTQVAQTTPHPTKLSRNYHPYIQTIDYSEPOLELITPI 70
 ++LP+L + G+ERE LR+ ++A TPHP+ LGS HP I TDYSE LE ITP
 Sbjct: 16 ANPLPLTECLHGIERECLRVSDSG-KLALTPHPRALGSTLTHTPQITTDYSEALLEFITPT 74
 30 Query: 71 AKDSQEAIRFLKAISDVAGRSINHDEYLWPLSMPPKV-REEDIQIAQLEDA----FEYDY 125
 D ++ L+ I A ++ EYLW SMP ++ EE I IA+ + +Y Y
 Sbjct: 75 ETDVADTLGDLEIRHFRASSKLD-GEYLWSPSPMPCPLPDEESIPIARYGSSMIGRLKVVY 133
 35 Query: 126 RKYLEKTYGKLIQSIGIHYNLGLGOELLTSLFELSQAD-NAIDFQNLQYMKLSQNFRLY 184
 RK L YGK +Q I+GIHYN L + L L + ++ + D+Q+ Y+ L +NF RY
 Sbjct: 134 RKGLALRYGKTMQCIAGIHYNFSPLPERLWPLLRLQAEGSELSESDYQSAAYIALIRNFRY 193
 40 Query: 185 RWLLTYLYGASPVAEEDFLDQKLNPNVR-----SLRNSHLGYVNHKDIRIS-- 230
 WLL YL+GASP + FL ++ R SLR S LGY N+ ++
 Sbjct: 194 SWLLMYLFGASPALDAGFLRGRPSQLERLDEHTLYLPYATSLRMSDLGYQNNAAQAGLTPC 253
 45 Query: 231 YTSLEKDYVNDLENV-----KSGQLIAEKEFYSPVRLR-----G 264
 Y L+ Y++ L AV + L E E+YS +R + G
 Sbjct: 254 YNDLQSYIDS LRQAVSTPYPPYKVGTKQDGEWVQLNTINILQIENEYSSIRPKRVITYTG 313
 50 Query: 265 SKACRNYLEKGITYLEFRTFDLNPFPISGITQETVDTVHFLFLLALLWIDS----- 314
 + + +G+ Y+E R D+NPF P+GI + + FLL + DS
 Sbjct: 314 ERPVQALAAARGVQYVEVRCLDINPFLPLGIDLDEARFLDAFLFCAFSDSPLLNGECSDA 373
 55 Query: 315 SSHIDQDIKEANRLN-DLIALSHPLEKLPNQAPVSDLVDMQSVIQHFNLSPPYQDLLES 373
 + + +KE R L P+E + + + +++ + L +
 Sbjct: 374 TDNFLAVVKEGRPGQLQRRGQPVQLQVWANELLERIADTAALLDRARGGEAHAALAA 433
 Query: 374 VKRQIQSPELTVAGQLLEMI--EGLSLETFGQRQGGIYHDYAWAEAPYA 419
 + ++ ELT + Q+L+++ G S E F RQ + + +Y + P A
 Sbjct: 434 QRAKVADAELTPSAQVLKVMRERGSEFAFSLRQSRHAHYFRQHPLA 481

-1639-

There is also homology to SEQ ID 4560.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1483

- 5 A DNA sequence (GBSx1569) was identified in *S.agalactiae* <SEQ ID 4561> which encodes the amino acid sequence <SEQ ID 4562>. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1504(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB73814 GB:AL139078 helix-turn-helix containing protein
[Campylobacter jejuni]
Identities = 107/223 (47%), Positives = 148/223 (65%), Gaps = 7/223 (3%)

20 Query: 1 MDKEKLDYWKTIITFLHNVLDGNYEIVLHVVDENDIYIGELVNSHISGRITSSPLTTFAL 60
MD+ + + FL VLG+ YEIV HV+ E+ YI + NSHISGR++ SPLT FA
Sbjct: 1 MDEGQKQOFIKLTYFLGEVLGEQYEIVFHVITEDGAYIAAIAANSHISGRSLDSPLTAFAS 60

25 Query: 61 DLIKNKVYKEKDFVTNYKAIVSPLNKEVRGSTFFIKNAQNELEGMLCINLDISAYQNIAL 120
+L++NK Y EKDF+ +YKA+V +K +RGSTFFIKN ++L G+LCIN D S +++
Sbjct: 61 ELMQNKKYLEKDFLCDYKALVGK-SKLIRGSTFFIKN-HDKLVGILCINHDTSIMRDLIC 118

Query: 121 DILDVLNL-NVNKILPKSPQKISLPQQEPEVLSGNIQDIISEIVDPSLLNQNIHLSQE 179
++DL + ++ IL IS Q + +E LS +I+DI+ + VD S LN + LS

30 Sbjct: 119 KMIDLEKIGDMGDIL----GNISFSQNDSSIETLSHSIEDILVQSVDSSYLNSDYQLSIT 174

Query: 180 VKVEIVSKLHEKGVFQLKGAVSKVAEVLNISEPSVYRYLKKIE 222
K EI KL+EKG+F +KGAV VA+ L ISEPSVYRYLKK +

35 Sbjct: 175 QKEEIAEKLYEKGIFPNKGAVPPIVAKFLKISEPSVYRYLKKFK 217

- 35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4563> which encodes the amino acid sequence <SEQ ID 4564>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1636(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 169/224 (75%), Positives = 198/224 (87%), Gaps = 3/224 (1%)

50 Query: 1 MDKEKLDYWKTIITFLHNVLDGNYEIVLHVVDENDIYIGELVNSHISGRITSSPLTTFAL 60
MDKE L+YWK+ITFLH+VLGDNYEI+LHV+D+NDIYIGELVNSHISGR+ SPLTTFAL
Sbjct: 1 MDKETLNYWKTVITFLHDVLGDNYEIIILHVIDKNDIYIGELVNSHISGRSKQSPLTTFAL 60

Query: 61 DLIKNKVYKEKDFVTNYKAIVSPLNKEVRGSTFFIKNAQNELEGMLCINLDISAYQNIAL 120
DLI NKVYKEKDFVTNYKAIVSP +KEVRGSTFFIK+ + LEGMLCINLDISAYQ +A

55 Sbjct: 61 DLITNKVYKEKDFVTNYKAIVSPQHEVRGSTFFIKDKKNLEGMLCINLDISAYQGVAR 120

Query: 121 DILDVLNLNVNKLIP--KSPQKISLPQQEPEVLSGNIQDIISEIVDPSLLNQNIHLSQ 178
D+L LVNLN+ +P K P+ ++ PQ EE VE+L+ NIQDII +I+DPSLL N+HLSQ

-1640-

Sbjct: 121 DLLKLVNLEHFIPTAKEPKIVT-PQPEEAVEILTSNIQDIIGQIIDPSLLRHNHLSQ 179

Query: 179 EVKVEIVSKLHEKGVFQLKGAVSKVAEVLNISEPSVYRYLKKIE 222
+VK++IV+KL+EKGVFQLKGAVSKVA++L ISEPSVYRYLKKIE

5 Sbjct: 180 DVKIDIVAKLYEKGVFQLKGAVSKVADILCISEPSVYRYLKKIE 223

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1484

10 A DNA sequence (GBSx1570) was identified in *S.galactiae* <SEQ ID 4565> which encodes the amino acid sequence <SEQ ID 4566>. This protein is predicted to be regulatory protein pfoR. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

15 INTEGRAL Likelihood = -7.80 Transmembrane 299 - 315 (296 - 325)
INTEGRAL Likelihood = -7.54 Transmembrane 172 - 188 (169 - 193)
INTEGRAL Likelihood = -7.17 Transmembrane 71 - 87 (66 - 98)
INTEGRAL Likelihood = -4.99 Transmembrane 261 - 277 (260 - 278)
20 INTEGRAL Likelihood = -2.81 Transmembrane 128 - 144 (127 - 149)
INTEGRAL Likelihood = -2.18 Transmembrane 101 - 117 (101 - 119)
INTEGRAL Likelihood = -0.53 Transmembrane 198 - 214 (197 - 214)

----- Final Results -----

25 bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA60239 GB:X86525 pfoS [Clostridium perfringens]

30 Identities = 96/147 (65%), Positives = 122/147 (82%)

Query: 100 GTGIIIPGFLAGYLVGFLVKWMERNIPGGGLDLISIIIGAPLTRLVAKLLTPLINSTLLTI 159
G GI+PGF+AGYL F++K++E+ IP GLDLI II++GAPL R +A + PL+ +TL I

35 Sbjct: 1 GFGILPGFIAGYLGsfVIKfLEKKIPAGLDLIVIVLGAPLVRGIAAISNPLVETTLQNI 60

Query: 160 GDILTSGAHSNPILMGIILGGTIVVVATAPLSSMALTAMLGLTGMPMAIGALSVFGSSFM 219
G ++T+ + ++PI+MGIILGG + VVATAPLSSMALTAMLGLTG+PMAIGAL+VFGSSFM

40 Sbjct: 61 GGVITATSTASPIIMGIILGGIVTVVATAPLSSMALTAMLGLTGLPMAIGALAVFGSSFM 120

Query: 220 NGVLPFKLKLGSRKDNIAFAVEPLTQA 246

N V F K+K GS+KD IA A+EPLTQA

Sbjct: 121 NLVFFGKMKFGSKKDTIAVAIEPLTQA 147

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4567> which encodes the amino acid sequence <SEQ ID 4568>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

50 INTEGRAL Likelihood = -8.70 Transmembrane 303 - 319 (296 - 325)
INTEGRAL Likelihood = -7.11 Transmembrane 70 - 86 (66 - 98)
INTEGRAL Likelihood = -6.53 Transmembrane 172 - 188 (169 - 193)
INTEGRAL Likelihood = -4.83 Transmembrane 261 - 277 (260 - 278)
INTEGRAL Likelihood = -2.55 Transmembrane 101 - 117 (101 - 119)
INTEGRAL Likelihood = -2.28 Transmembrane 124 - 140 (124 - 140)
55 INTEGRAL Likelihood = -1.91 Transmembrane 198 - 214 (197 - 215)

----- Final Results -----

bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1641-

The protein has homology with the following sequences in the databases:

>GP:CAA60239 GB:X86525 pFoS [Clostridium perfringens]
Identities = 95/147 (64%), Positives = 123/147 (83%)

5 Query: 100 GTGIIPGFVAGYVVSFLIKWMEKNIPGGDLISIIIVGAPLTRFLAQLITPVINSTLLTI 159
G GI+PGF+AGY+ SF+IK++EK IP GLDLI II++GAPL R +A + P++ +TL I
Sbjct: 1 GFGILPGFIAGYLGSSFVIKFKLEKKIPAGLDLIVIVLGAPLVRGIAAISNPLVETTLQNI 60

10 Query: 160 GDILTSSANSNPIIMGMILGGTIVVVATAPLSSMALTAMLGLTGIPMAIGALSVFGSSFM 219
G ++T+++ ++PI+MG+ILGG + VVATAPLSSMALTAMLGLTG+PMAIGAL+VFGSSFM
Sbjct: 61 GGVITATSTASPIIMGGIILGGIVTVVATAPLSSMALTAMLGLTGIPMAIGALAVFGSSFM 120

15 Query: 220 NGVLFYRLKLGKRDNI AFAIEPLTQA 246
N V F ++K G +KD IA AIEPLTQA
Sbjct: 121 NLVFFGKMKFGSKKDTIAVAIEPLTQA 147

An alignment of the GAS and GBS proteins is shown below.

Identities = 302/339 (89%), Positives = 330/339 (97%)

20 Query: 1 MNIIIGTSLILVLAIFTLFNYKAPYGTKAMGALASAACASFLVEAFQDSFFGKVLGFQF 60
M+IIIGTSLILVLAI F+LFNYKAP+G KAMGALASAACASFLVEAFQDSFFGKVLGFQF
Sbjct: 1 MDIIIGTSLILVLAI FSLFNYKAPHGAKAMGALASAACASFLVEAFQDSFFGKVLGFQF 60

25 Query: 61 LSEVGGANGSLSGVAAAILVAIAIGVTPGYAVLIGLSVSGTGIIPGFLAGYLVGFLVKWM 120
LSEVGGANGSLSGVAAAILVAIAIGV+PGYAVLIGLSVSGTGIIPGF+AGY+V FL+KWM
Sbjct: 61 LSEVGGANGSLSGVAAAILVAIAIGVSPGYAVLIGLSVSGTGIIPGFVAGYVVSFLIKWM 120

30 Query: 121 ERNIPGGDLISIIIGAPLTRLVAKLLTPLINSTLLTIGDILTSGAHSNPILMGIILGG 180
E+NIPGGDLISIII+GAPLTR +A+L+TP+INSTLLTIGDILTS A+SNPI+MG+ILGG
Sbjct: 121 EKNIPGGDLISIIIVGAPLTRFLAQLITPVINSTLLTIGDILTSSANSNPIIMGMILGG 180

35 Query: 181 TIVVVATAPLSSMALTAMLGLTGMPMAIGALSVFGSSFMNGVLFHKLKLSRKDNIAFAV 240
TIVVVATAPLSSMALTAMLGLTG+PMAIGALSVFGSSFMNGVLF++LKLK RKDNIAFA+
Sbjct: 181 TIVVVATAPLSSMALTAMLGLTGIPMAIGALSVFGSSFMNGVLFYRLKLGKRDNI AFAI 240

40 Query: 241 EPLTQADVTSANPIPIYVTNFVGGAAACGILIALMKLVNDTPGTATPIAGFAVMFAYNPNI 300
EPLTQADVTSANPIPIYVTNFVGGAAACG+LIALMKLVNDTPGTATPIAGFAVMFAYNP+
Sbjct: 241 EPLTQADVTSANPIPIYVTNFVGGAAACGVLI ALMKLVNDTPGTATPIAGFAVMFAYNPVA 300

45 Query: 301 KVLITALGCIISLLAGYFGGIVFKDYKLVTKKEELQARD 339
KVLITALGCI+SL+ GY GG VFK+Y+LVTK+ELQAR+
Sbjct: 301 KVLITALGCIISLIVGYIGGSVFKNYRLVTKQELQARN 339

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1485

A DNA sequence (GBSx1571) was identified in *S.agalactiae* <SEQ ID 4569> which encodes the amino acid sequence <SEQ ID 4570>. This protein is predicted to be adenylosuccinate synthetase (purA). Analysis of this protein sequence reveals the following:

Possible site: 27
>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0560 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

-1642-

>GP:CAB16079 GB:Z99124 adenylosuccinate synthetase [Bacillus subtilis]
Identities = 320/427 (74%), Positives = 378/427 (87%)

5 Query: 1 MTSVVVGTQWGDEGKGKITDFLSADAEVIARYQGGDNAGHTIVIDNKKFKLHLPISGIF 60
M+SVVVVGTQWGDEGKGKITDFLS +AEVIARYQGG+NAGHTI D +KLHLPISGIF
Sbjct: 1 MSSVVVGTQWGDEGKGKITDFLSENAEVIARYQGGNNAGHTIKFDGITYKLHLPISGIF 60

10 Query: 61 FKEKISVIGNGVVNPVSLVKELAYLHGEVTTDNLRIISDRAHVILPYHIKLDQLQEDAK 120
+K+K VIGNG+VV+PK+LV ELAYLH V+TDNLRIIS+RAHVILPYH+KLD+++E+ K
Sbjct: 61 YKDKTCVIGNGMVDPKALVTELAYLHERNVSTDNLRIISRAHVILPYHLKLDEVEEERK 120

15 Query: 121 GDNKIGTTIKGIGPAYMDKAARVGIRIADLLDREVFAERLKNLAEKNRLFEEKMYDSTPL 180
G NKIGTT KGIGPAYMDKAAR+GIRIADLLDR+ FAE+L+ NL EKNRL EKM++
Sbjct: 121 GANKIGTTKKGIGPAYMDKAARIGIRIADLLDRDAFAEKLERNLEKNRLLEKMYETEGF 180

20 Query: 181 EFDDIFEYEEYEGQKIKQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS 240
+ +DI +EYEEYEGQKIK+YV DTSV+LNDALD G+RVLFEGAQGVMLDIDQGTYPFVTSS
Sbjct: 181 KLEDILDEYEEYEGQKIKKYVCDTSVVLNDALDEGRVLFEGAQGVMLDIDQGTYPFVTSS 240

25 Query: 241 NPVAGGVTIGSGVGPSKINKVVGVCAYTSRVGDGPFPTLDFEVGDRIREIGKEYGTTT 300
NPVAGGVTIGSGVGP+KI VVGVCAYT+RVGDGPFPTL DE+GD+IRE+G+EYGTIT
Sbjct: 241 NPVAGGVTIGSGVGP+KIKHVVGVCAYTTRVGDGPFPTLDEIGDQIREVGREYGTIT 300

30 Query: 301 GRPRRVGWFDVSVMRHSRRVSGITNLSLNSIDVLSGLDVKICVAYDLGKRIDYYPASL 360
GRPRRVGWFDVSV+RH+RRVSGIT+LSLNSIDVL+G++T+KICVAY G+ I+ +PASL
Sbjct: 301 GRPRRVGWFDVSVVRHARRVSGITDLSLNSIDVLGAGTETLKICVAYRYKGEIIEFPASL 360

35 Query: 361 EQLKRCKPIYEELPGWSEDTACRSIDDLPENARNYVRRVGVGVRISTFVSGPGRQET 420
+ L C+P+YEE+PGW+EDIT +SL +LPENAR+Y+ RV +L G+ +S FSVGP R QT
Sbjct: 361 KALAECEPVYEEEMPWTEDITGAKSLSELPENARHYLERVSQITGIPLSIFSVGPDRSQT 420

40 Query: 421 NILESVM 427
N+L SV+
Sbjct: 421 NVLRSVY 427

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4571> which encodes the amino acid sequence <SEQ ID 4572>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0560(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 406/430 (94%), Positives = 421/430 (97%)

50 Query: 1 MTSVVVGTQWGDEGKGKITDFLSADAEVIARYQGGDNAGHTIVIDNKKFKLHLPISGIF 60
MTSVVVVGTQWGDEGKGKITDFLSADAEVIARYQGGDNAGHTIVID KKFKLHLPISGIF
Sbjct: 1 MTSVVVGTQWGDEGKGKITDFLSADAEVIARYQGGDNAGHTIVIDGKKFKLHLPISGIF 60

55 Query: 61 FKEKISVIGNGVVNPVSLVKELAYLHGEVTTDNLRIISDRAHVILPYHIKLDQLQEDAK 120
F +KISVIGNGVVNPVSLVKELAYLH EGVTTDNLRIISDRAHVILPYHI+LDQLQEDAK
Sbjct: 61 FPQKISVIGNGVVNPVSLVKELAYLHDEGVTTDNLRIISDRAHVILPYHIQLDQLQEDAK 120

60 Query: 121 GDNKIGTTIKGIGPAYMDKAARVGIRIADLLDREVFAERLKNLAEKNRLFEEKMYDSTPL 180
GDNKIGTTIKGIGPAYMDKAARVGIRIADLLD+++FAERL+INLAEKNRLFEEKMYDSTPL
Sbjct: 121 GDNKIGTTIKGIGPAYMDKAARVGIRIADLLDKDIFAERLINLAEKNRLFEEKMYDSTPL 180

65 Query: 181 EFDDIFEYEEYEGQKIKQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS 240
+FD IFEEY YGQ+IKQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS
Sbjct: 181 DFDAIFEYEEYAYGQKIKQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS 240

Query: 241 NPVAGGVTIGSGVGPSKINKVVGVCAYTSRVGDGPFPTLDFEVGDRIREIGKEYGTTT 300

-1643-

NPVAGGVTTIGSGVGP+KINKVVGVCAYTSRVGDGPFPTLFDVVG+RIRE+G EYGTTT
 Sbjct: 241 NPVAGGVTTIGSGVGNKINKVVGVCAYTSRVGDGPFPTLFDVVGIRREVGEYGTIT 300

Query: 301 GRPRRVGWFDVVMRHSRRVSGITNLSLNSIDVLSGLDVTVKICVAYDLGKRIDYYPASL 360
 GRPRRVGWFDVVMRHSRRVSGITNLSLNSIDVLSGLDVTVKICVAYDLGKRIDYYPAL 360
 Sbjct: 301 GRPRRVGWFDVVMRHSRRVSGITNLSLNSIDVLSGLDVTVKICVAYDLGKRIDYYPANL 360

Query: 361 EQLKRCKPIYEELPGWSEDTACRSLDDLPENARNYVRRVGLVGVRISTFSVGPGRQET 420
 EQLKRCKPIYEELPGW EDIT RSLD+LPENARNYVRRVGLVGVRISTFSVGPGRQET
 Sbjct: 361 EQLKRCKPIYEELPGWQEDITGVRSLDELPENARNYVRRVGLVGVRISTFSVGPGRQET 420

Query: 421 NILESVMWSNI 430
 NILESVM++I
 Sbjct: 421 NILESVMWASI 430

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1486

A DNA sequence (GBSx1572) was identified in *S.agalactiae* <SEQ ID 4573> which encodes the amino acid sequence <SEQ ID 4574>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.29	Transmembrane	30 - 46 (22 - 55)
INTEGRAL	Likelihood = -2.97	Transmembrane	110 - 126 (109 - 126)
INTEGRAL	Likelihood = -0.11	Transmembrane	89 - 105 (89 - 106)

----- Final Results -----
 bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8823> which encodes amino acid sequence <SEQ ID 8824> was also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 10
 SRCFLG: 0
 McG: Length of UR: 5
 Peak Value of UR: 3.05
 Net Charge of CR: 0
 McG: Discrim Score: 4.64
 GvH: Signal Score (-7.5): -1.66
 Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.
 Amino Acid Composition: calculated from 37
 ALOM program count: 2 value: -2.97 threshold: 0.0

INTEGRAL	Likelihood = -2.97	Transmembrane	100 - 116 (99 - 116)
PERIPHERAL	Likelihood = 1.38		56

modified ALOM score: 1.09
 icml HYPID: 7 CFP: 0.219

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.2190(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database and no corresponding DNA sequence was identified in *S.pyogenes*.

-1644-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1487

A DNA sequence (GBSx1573) was identified in *S.agalactiae* <SEQ ID 4575> which encodes the amino acid sequence <SEQ ID 4576>. Analysis of this protein sequence reveals the following:

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.0967(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1488

A DNA sequence (GBSx1574) was identified in *S.agalactiae* <SEQ ID 4577> which encodes the amino acid sequence <SEQ ID 4578>. This protein is predicted to be SgaT protein (sgaT). Analysis of this protein sequence reveals the following:

Possible site: 43
>>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -7.80 Transmembrane 441 - 457 (436 - 464)
 INTEGRAL Likelihood = -7.64 Transmembrane 344 - 360 (339 - 376)
 INTEGRAL Likelihood = -6.58 Transmembrane 403 - 419 (392 - 422)
 INTEGRAL Likelihood = -6.48 Transmembrane 237 - 253 (235 - 261)
 INTEGRAL Likelihood = -5.79 Transmembrane 105 - 121 (99 - 127)
30 INTEGRAL Likelihood = -5.52 Transmembrane 138 - 154 (137 - 155)
 INTEGRAL Likelihood = -4.78 Transmembrane 18 - 34 (14 - 38)
 INTEGRAL Likelihood = -2.97 Transmembrane 365 - 381 (365 - 383)
 INTEGRAL Likelihood = -0.69 Transmembrane 41 - 57 (41 - 57)
 INTEGRAL Likelihood = -0.16 Transmembrane 160 - 176 (160 - 176)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77150 GB:AE000491 crf, hypothetical protein [Escherichia coli K12]
Identities = 181/451 (40%), Positives = 274/451 (60%), Gaps = 25/451 (5%)

45 Query: 11 FSQNILQNPAFFVGLLVLLIGYLLKKPLHDVFAGFIKATVGYLILNVGAGGLVNTFRPIL 70
 F ++ N +G++ +GY+LL+K + + G IK +G+++L G+G L +TF+P++
Sbjct: 30 FFNQVMTNAPLLLGIVTCLGYILLRKSVSVIIKGTIKTIIGFMLLQAGSGILITSTFKPVV 89

Query: 71 VALAKKFNLEAAVIDPYFGLASANAKLETMG-FISVATTALLIGFGINILLVALRKVTKV 129
 +++ + + A+ D Y AS A ++ MG S A+L+ +NI V LR++T +
50 Sbjct: 90 AKMSEVYGINGAISDTY---ASMMATIDRMGDAYSWSVGYAVLLALALNICYVLLRRITGI 146

Query: 130 RTLFI TGHIMVQQAATISVFVLLIPQLRNGFGAWAV----GIICGLYWAVSSNMTVEAT 185
 RT+ +TGHIM QQA I+V + + G+ W I+ LYW ++SNM + T
Sbjct: 147 RTIMLTGHIMFQQAGLIAVTLFIF-----GYSMWTTTICTAILVSLYWGITSNMMYKPT 200

-1645-

Query: 186 QRLTGGGGFAIGHQQQFAIWFVDKVPFFGKKEENLDNLKLPFLNIFHDTVVASATLML 245
 Q +T G GF+IGHQQQFA W KVAPF GKKEE++++LKL +LNIFHD +V++A +M
 Sbjct: 201 QEVTDGCGFSIGHQQQFASWIAVKVAPFLGKKEESVEDLKLPGWLNIFHDNIVSTAIVMT 260

Query: 246 VFFGGILAVLPGDIMSNNKLGPGAFVPTKQAFMYILQTSLTFSVYLFILMQGVRMFVT 305
 +FFG IL G D + + K + +YILQT +F+V +FI+ QGVRMFV
 Sbjct: 261 IFFGAILLSFGIDTVQ-----AMAGKVHWTVYILQTSFSAVAIFITQGVRMFVA 311

Query: 306 ELTNAFQGISNKLPGSFPDVAASYGFGSSNAVLSGFAGLIGQLITIALLVVFNPI 365
 EL+ AF GIS +L+PG+ A+D AA Y F + NAV+ GF +G IGQLI + +LV + I
 Sbjct: 312 ELSEAFNGISQRLIPGAVLAIDCAIYSF-APNAVVGFMWGTIGQLIAVGILVACGSSI 370

Query: 366 LIITGFVPVFFDNAAIAVYADKRGWKAVALSFISGIIQVALGAVAVGLLGLAGGYHGN 425
 LII GF+P+FF NA I V+A+ GGW+AA+ + + G+I++ AV L G++ + G
 Sbjct: 371 LIIPGFIPMFFSNATIGVFANHFGGWRAALKICLVGMIEIFGCVWAVKLTGMS-AWMGM 429

Query: 426 IDFEFPWLAFGYIFKYLGIAGYVIVCLFFLA 456
 D+ F +GIA ++ + LA
 Sbjct: 430 ADWSILAPPMQGFSGIAFMAVIVIALA 460

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4579> which encodes the amino acid sequence <SEQ ID 4580>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.51	Transmembrane	441 - 457 (435 - 465)
INTEGRAL	Likelihood = -7.80	Transmembrane	344 - 360 (339 - 376)
INTEGRAL	Likelihood = -7.64	Transmembrane	238 - 254 (235 - 261)
INTEGRAL	Likelihood = -5.63	Transmembrane	105 - 121 (100 - 127)
INTEGRAL	Likelihood = -5.52	Transmembrane	138 - 154 (137 - 155)
INTEGRAL	Likelihood = -5.20	Transmembrane	400 - 416 (392 - 422)
INTEGRAL	Likelihood = -4.78	Transmembrane	18 - 34 (14 - 39)
INTEGRAL	Likelihood = -2.97	Transmembrane	365 - 381 (365 - 383)
INTEGRAL	Likelihood = -1.49	Transmembrane	160 - 176 (160 - 177)
INTEGRAL	Likelihood = -0.53	Transmembrane	41 - 57 (41 - 57)

----- Final Results -----
 bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC77150 GB:AE000491 orf, hypothetical protein [Escherichia coli]
 Identities = 182/461 (39%), Positives = 279/461 (60%), Gaps = 25/461 (5%)

Query: 1 MEMLLAPLNWFSQNLQNPFAFFVGLLVLLIGYLLKKPIYEVFAGFKATVGYLILNVGAG 60
 ME+L F ++ N +G++ +GY+LL+K + + G +K +G+++L G+G
 Sbjct: 20 MEILYNIFTVFFNQVMTNAPLLGLIVTCLGYILLRKSVSVIIRKTIKTIIGFMLLQAGSG 79

Query: 61 GLVITFRPILVALAKKFELKAAVIDPYFGLAAANTKLEEMG-FISVATTALLIGFGVNIL 119
 L +TF+P++ +++ + + A+ D Y + A ++ MG S A+L+ +NI
 Sbjct: 80 ILTSTFKPVVAKMSEVYINGAISDTYASMMAT---IDRMGDAYSWVGAVLLALALNIC 136

Query: 120 LVALRKVTQVRLTFTIGHIMVQQAATISVFVLLIPQFQNAFGAWAV---GIICGLYWA 175
 V LR++T +RT+ +TGHIM QQA I+V + + + W I+ LYW
 Sbjct: 137 YVLLRRITGIRTIMLTGHIMFQQAGLIAVTLFIF-----GYSMWTTIICTAILVSLYWG 190

Query: 176 ISSNMTVEATQRLTGGGGFAIGHQQQFAIWFVDKVPFFGKKEENLDNLKLPFLNIFHD 235
 I+SNM + TQ +T G GF+IGHQQQFA W KVAPF GKKEE++++LKL +LNIFHD
 Sbjct: 191 ITSNNMYKPTQEVTDGCGFSIGHQQQFASWIAVKVAPFLGKKEESVEDLKLPGWLNIFHD 250

Query: 236 TVVASATLMLVFFGAILAVLPGDIMSDDVLIGPGAFNPAKQAFMYILQTSLTFSVYLF 295
 +V++A +M +FFGAIL G D + + K + +YILQT +F+V +FI
 Sbjct: 251 NIVSTAIVMTIFFGAILLSFGIDTVQAM-----AGKVHWTYVILQTSFSAVAIFI 301

-1646-

Query: 296 LMQGVRMFVSELTNAFQGISKLLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITI 355
 + QGVRMFV+EL+ AF GIS +L+PG+ A+D AA Y F + NAV+ GF +G IGQLI +
 Sbjct: 302 ITQGVRMFVAELSEAFNGISQRLIPGAVLAIDCAAISYF-APNAVVGFMWGTIGQLIAV 360

Query: 356 ALLVIFKNPILIITGFVPVFFDNAAIAVYADKRGGWKAVALSFISGILQVALGAVAVGL 415
 +LV + ILII GF+P+FF NA I V+A+ GGW+AA+ + + G++++ AV L
 Sbjct: 361 GILVACGSSILIIPGFIPMFSSNATIGVFANHFQGWRAALKICLVMGMIIEIFGCVWAVKL 420

Query: 416 LGLTGGYHGNIDLVLPLPFGYLFKFLGIAGYVLCIFLLA 456
 G++ + G D + P F +GIA ++ + LA
 Sbjct: 421 TGMS-AWMMGADWSILAPPMQGFPSIGIAFMAVITVIALA 460

An alignment of the GAS and GBS proteins is shown below.

Identities = 437/476 (91%), Positives = 457/476 (95%)

Query: 1 MENFLAPLNWFSQNILQNPAFFVGLLVLLIGYLLKKPLHDVFAGFIKATVGYLILNVGAG 60
 ME LAPLNWFSQNILQNPAFFVGLLVLLIGYLLKKP+++VFAGF+KATVGYLILNVGAG
 Sbjct: 1 MEMLLAPLNWFSQNILQNPAFFVGLLVLLIGYLLKKPIYEVFAGFVKATVGYLILNVGAG 60

Query: 61 GLVNTFRPILVALAKKFLEAAVIDPYFGLASANAKLETMGFISVATTALLIGFGINILL 120
 GLV TFRPILVALAKKF L+AAVIDPYFGLA+AN KLE MGFISVATTALLIGFG+NILL
 Sbjct: 61 GLVTTFRPILVALAKKFELKAAVIDPYFGLAAANTKLEEMGFISVATTALLIGFGVNILL 120

Query: 121 VALRKVTKVRTLFTITGHIMVQQAATISVFVLLIPQLRNGFGAWAVGIICGLYWAVSSNM 180
 VALRKVTKVRTLFTITGHIMVQQAATISVFVLLIPQ +N FGAWAVGIICGLYWA+SSNM
 Sbjct: 121 VALRKVTKVRTLFTITGHIMVQQAATISVFVLLIPQFQNAFGAWAVGIICGLYWAISNM 180

Query: 181 TVEATQRLTGGGGFAIGHQQQFAIWFVDKVAPFFGKKEENLDNLKLPFTFLNIFHDTVVAS 240
 TVEATQRLTGGGGFAIGHQQQFAIWFVDKVAPFFGKKEENLDNLKLPFTFLNIFHDTVVAS
 Sbjct: 181 TVEATQRLTGGGGFAIGHQQQFAIWFVDKVAPFFGKKEENLDNLKLPFTFLNIFHDTVVAS 240

Query: 241 ATLMVLVFFGGILAVLGPDIMSNVKLIGPGAFVPTKQAFFMYILQTSLTFSVYLFILMQGV 300
 ATLMVLVFFG ILAVLGPDIMSV LIGPGAF P KQAFFMYILQTSLTFSVYLFILMQGV
 Sbjct: 241 ATLMVLVFFGAILAVLGPDIMSDVDLIGPGAFNPAKQAFFMYILQTSLTFSVYLFILMQGV 300

Query: 301 RMFVTELTNAFQGISNKLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITIALLVV 360
 RMFV+ELTNFQGIS+KLLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITIALLV+
 Sbjct: 301 RMFVSELTNAFQGISKLLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITIALLVI 360

Query: 361 FKNPILIITGFVPVFFDNAAIAVYADKRGGWKAVALSFISGIIQVALGAVAVGLLGLAG 420
 FKNPILIITGFVPVFFDNAAIAVYADKRGGWKAVALSFISGI+QVALGAVAVGLLGL G
 Sbjct: 361 FKNPILIITGFVPVFFDNAAIAVYADKRGGWKAVALSFISGILQVALGAVAVGLLGLTG 420

Query: 421 GYHGNIDFEFPWLAFGYIFKYLGIAGYVIVCLFFLAIPQLQFMKSKDKEAYYRGDA 476
 GYHGNID PWL FGY+FK+LGIAGYV+VC+F LAIPQLQF K+KDKEAYYRG+A
 Sbjct: 421 GYHGNIDLVLPLPFGYLFKFLGIAGYVLCIFLLAIPQLQFAKAKDKEAYYRGEA 476

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1489

A DNA sequence (GBSx1575) was identified in *S. galactiae* <SEQ ID 4581> which encodes the amino acid sequence <SEQ ID 4582>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1225 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-1647-

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG34743 GB:AE000033 similar to PTS system: EIIB [Mycoplasma pneumoniae]
Identities = 40/89 (44%), Positives = 62/89 (68%), Gaps = 1/89 (1%)

5 Query: 4 VLTACGNGMGSSMVIKMKVENALRQLGVSNFESASCSVGEAKGLAANYDIVVASNHLIHE 63
++ ACGNGMG+SM+IK+KVE +++LG + A S+G+ KG+ + DI+++S HL E
Sbjct: 8 IIAACGNGMGTSMLIKIKVEKIMKELGYTAKVEA-LSMGQTKGMEHSADIITISSIHLTSE 66
10 Query: 64 LDGRTKGHLVGLDNLMDNEIKTKLQEIL 92
+ K +VG+ NLMD+NEIK L ++L
Sbjct: 67 FNPNAKAKIVGVNLNLMDENEIKQALSKVL 95

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4583> which encodes the amino acid sequence <SEQ ID 4584>. Analysis of this protein sequence reveals the following:

15 Possible site: 42
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
20 bacterial cytoplasm --- Certainty=0.0977(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 85/92 (92%), Positives = 90/92 (97%)
25 Query: 1 MVKVLTAACGNGMGSSMVIKMKVENALRQLGVSNFESASCSVGEAKGLAANYDIVVASNHL 60
MVKVLTAACGNGMGSSMVIKMKVENALRQLGV++ +SASCSVGEAKGLA+ YDIVVASNHL
Sbjct: 1 MVKVLTAACGNGMGSSMVIKMKVENALRQLGVTDIQSASCSVGEAKGLASGYDIVVASNHL 60
30 Query: 61 IHELDGRTKGHLVGLDNLMDNEIKTKLQEIL 92
IHELDGRTKGHLVGLDNLMDNEIKTKLQE+L
Sbjct: 61 IHELDGRTKGHLVGLDNLMDNEIKTKLQEV L 92

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
35 vaccines or diagnostics.

Example 1490

A DNA sequence (GBSx1576) was identified in *S.agalactiae* <SEQ ID 4585> which encodes the amino acid sequence <SEQ ID 4586>. This protein is predicted to be a pentitol phosphotransferase enzyme ii, a component (ptxA). Analysis of this protein sequence reveals the following:

40 Possible site: 38
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
45 bacterial cytoplasm --- Certainty=0.3309(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77152 GB:AE000491 putative PTS system enzyme II A component
[Escherichia coli K12]
50 Identities = 64/150 (42%), Positives = 97/150 (64%), Gaps = 2/150 (1%)
Query: 1 MNLKQAFIENDSIRLKLSDWKEAIKLSIDPLIESGAVDAEYYDAIESTEEFGPYIYL 60
M L+ + EN SIRL+ A W+EA+K+ +D L+ + V+ YY AI++ E+FGPY+++
55 Sbjct: 1 MKLRDSLAKENKIRLQAEAEVQEAQVIGVLLVAADVPEPRYYQAILDGVEQFGPYFVI 60
Query: 61 MFGMAMPHARPEAGVKRDAFSLITLTPEVVF--PDGKEVSVLLALAATSSAIHTSVAIPQ 118

-1648-

PG+AMPH RPE GVK+ FSL+TL +P+ F D V +L+ +AA + H V I Q
 Sbjct: 61 APGLAMPHGRPEEGVKKTGFSLVTLKKPLEFNHDDNDPVDILITMAAVDANTHQEVGIMQ 120

Query: 119 IIALFELENSIQRLTECQEAKEVLAMVEES 148

I+ LFE E + RL C+ +EVL +++ +

Sbjct: 121 IVNLFEEDEENFDRLRACRTEQEVLDLIDRT 150

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4587> which encodes the amino acid sequence <SEQ ID 4588>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2287(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 113/161 (70%), Positives = 137/161 (84%)

Query: 1 MNLKQAFIENDSIRLKLASDWKRAIKLSIDPLIESGAVDAEYYDAIIESTEEFGPYIYL 60

MNLKQAFI+N+SIRL LSA W+EA++L++ PLI+S AV + YYDAII STE++GPYY+L

Sbjct: 1 MNLKQAFIDNNSIRLGLSADTWQEAURLAVQPLIDSKAVTSAYYDAIIASTEKYGPYYVL 60

Query: 61 MPGMAMPHARPEAGVKRDAFSLITLITPVPFDPGKEVSVLLALAAATSSAIHTSVAIPQII 120

MPGMAMPHA GV R+AF+LITLT+PV F DGKEVSVLL LAAT +IHT+VAIPQI+

Sbjct: 61 MPGMAMPHAEAGLGVNRNFAFALITLTKPVTFSKGKEVSVLLTLAATDPSIHTTVAIPQIV 120

Query: 121 ALFELENSIQRLTECQEAKEVLAMVEESKNSPYLEGLDLES 161

ALFEL+N+I+RL CQ KEVL MVEESK+SPYLEG+DL +

Sbjct: 121 ALFELDNAIERLVACQSPKEVLEMVEESKDSPYLEGMDLNA 161

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1491

A DNA sequence (GBSx1577) was identified in *S.galactiae* <SEQ ID 4589> which encodes the amino acid sequence <SEQ ID 4590>. This protein is predicted to be probable hexulose-6-phosphate synthase. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1584(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77153 GB:AE000491 probable hexulose-6-phosphate synthase
 [Escherichia coli K12]

Identities = 108/217 (49%), Positives = 141/217 (64%), Gaps = 3/217 (1%)

Query: 5 LPNLQVALDHSDDLQGAIKAAVSVGHEVDVIEAGTVCLLQVGSELVEVLRSLFPDKIIVAD 64

LP LQVALD+ + A + + EVD+IE GT+ + G V L++L+P KI++AD

Sbjct: 3 LPMLQVALDNQTMDSAYETTRLIAEEVDIIEVGTILCVGEGVRAVRDLKALYPHKIVLAD 62

Query: 65 TKCADAGGTVAKNNAVRGADWMTICCATIPTMEALKAIKEERGDRGEIQIELYGDWTY 124

K ADAG +++ ADW+T ICCA I T + AL KE GD +QIEL G WT+

Sbjct: 63 AKIADAGKILSRMCFEANADWVTVICCADINTAKGALDVAKFNGD---VQIELTGYWTW 119

-1649-

Query: 125 EQAQQWLDAGISQAIYHQSRDALLAGETWGEKDLNKVKLLIDMGFRVSVTGGLSTDITLQL 184
 EQAQQW DAGI Q +YH+SRDA AG WGE D+ +K+L DMGF+V+VTGGL+ + L L
 Sbjct: 120 EQAQQWRDAGIGQVVYHRSRDAQAAGVANGGEADITAIKRLSDMGFKVTVTGGLALEDLPL 179

Query: 185 FEGVDVFTFIAGRGITEADDPAAAAARAFKDEIKRIWG 221
 F+G+ + FIAGR I +A P AAR FK I +WG
 Sbjct: 180 FKGIPIHVFIAGRSIRDASPVEAARQFKRSIAELWG 216

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4591> which encodes the amino acid sequence <SEQ ID 4592>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1473 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 206/217 (94%), Positives = 212/217 (96%)

- Query: 5 LPNLQVALDHSDLQGAIKAAVSVGHEVDVIEAGTVCLLQVGSELVEVLRSLFDPDKIIVAD 64
 +PNLQVALDHSDLQGA+KAAV+VGHEVDVIEAGTVCLLQVGSELVEVLRSLFP+KIIVAD
 25 Sbjct: 4 IPNLQVALDHSDLQGAIVKAAVAVGHEVDVIEAGTVCLLQVGSELVEVLRSLFPEKIIVAD 63
- Query: 65 TKCADAGGTIVAKNNAVRGADWMTICCATIPTMEALKAIKEERGDRGEIQIELYGDWTY 124
 TKCADAGGTIVAKNNA RGADWMTICCATIPTMEALKAIKEERGDRGEIQIELYGDWTY
 30 Sbjct: 64 TKCADAGGTIVAKNNAKRGADWMTICCATIPTMEALKAIKEERGDRGEIQIELYGDWTY 123
- Query: 125 EQAQQWLDAGISQAIYHQSRDALLAGETWGEKDLNKVKLLIDMGFRVSVTGGLSTDITLQL 184
 EQAQ WLDAGISQAIYHQSRDALLAGETWGEKDLNKVK LIDMGFRVSVTGGL DTL+L
 Sbjct: 124 EQAQLWLDAGISQAIYHQSRDALLAGETWGEKDLNKVKTLLIDMGFRVSVTGGLDVTDLRL 183
- 35 Query: 185 FEGVDVFTFIAGRGITEADDPAAAAARAFKDEIKRIWG 221
 FEGVDVFTFIAGRGITEA+DPAAAAARAFKDEIKRIWG
 Sbjct: 184 FEGVDVFTFIAGRGITEADDPAAAAARAFKDEIKRIWG 220

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 40 vaccines or diagnostics.

Example 1492

A DNA sequence (GBSx1578) was identified in *S.galactiae* <SEQ ID 4593> which encodes the amino acid sequence <SEQ ID 4594>. Analysis of this protein sequence reveals the following:

- Possible site: 36
 45 >>> Seems to have no N-terminal signal sequence
- Final Results -----
 bacterial cytoplasm --- Certainty=0.4179 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- >GP:AAC22686 GB:U32783 hexulose-6-phosphate isomerase, putative
 [Haemophilus influenzae Rd]
 55 Identities = 143/282 (50%), Positives = 199/282 (69%), Gaps = 3/282 (1%)
- Query: 5 IGIYEKATPKHFNWLERLQPAKELGFDVFLSIDESDERLARLEWSKEERLELVKAIFET 64
 IGIYEKA PK+ W ERL AK GF+F+E+SIDES++RL+RL W+K ER+ L ++I ++
 Sbjct: 6 IGIYEKALPKNITWQERLSLAKACGFEFIEMSIDESNDRLSRLNWTKSERIALHQSIQS 65

-1650-

Query: 65 GVRVPTITFSGHRRFPMGSNNPEKEARAMMMKKCIVFAQDIGIRNIQLAGYDVYYEEKS 124
 G+ +P++ S HRRFP GS + + ++ ++M+K I + ++GIR IQLAGYDVYYE++
 Sbjct: 66 GITIPSMCLSAHRRFPFGSKDKIRQKSFEIMEKAIDLSVNLGIRTIQLAGYDVYYEKQD 125

Query: 125 PETRARFIKNLRQACTWAEAAQVILSIEIMDDPFMNSIEKYLAKEIDSPYLFVYPDTG 184
 ET F + + A T A AOV L++EIMD PFM+SI ++ + I+SP+ VYPD G
 Sbjct: 126 EETIKYFQEGIEFAVTLAASAQVTLAVEIMDTPFMSSISRWKKWDTIINSPWFTVYPDIG 185

Query: 185 NVSAWHNDLWSEFYNGHRSIAALHIKDTYAVTETSKGQFRDVPFGQGCVDWEEMFAVIKK 244
 N+SAW+N++ E G I+A+H+KDTY VTETSKGQFRDVPFGQGCVD+ F+++KK
 Sbjct: 186 NLSAWNNNIEBELTLGIDKISAIHLKDTYPVTETSKGQFRDVPFGQGCVDVHFFSLLKK 245

Query: 245 TNYNGPFLIEMWSENCETVEETRAAIKEAQDFLYPLMEKTGV 286
 NY G FLIEMW+E EE I +A+ ++ MEK G+
 Sbjct: 246 LNYRGAFLEIEMWTEK---NEEPLLEIIQARKWIVQQMEKAGL 284

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4595> which encodes the amino acid sequence <SEQ ID 4596>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1489(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 240/286 (83%), Positives = 271/286 (93%)

Query: 1 MTRPIGIYEKATPKHFNWLERLQFAKELGFDVELSIDESDERLARLEWSKEERLELVKA 60
 M RPIGIYEKATPK F W ERLQPAK+LGFDVE+S+DESD RLARLEW+KEERL+LVKA
 Sbjct: 15 MARPIGIYEKATPKQFTWRERLQFAKDLGFDVEMSVDESARLARLEWTKEERLDLVKA 74

Query: 61 IFETGVRVPTITFSGHRRFPMGSNNPEKEARAMMMKKCIVFAQDIGIRNIQLAGYDVYY 120
 I+ETG+R+PTI FSGHRR+P+GSN+P EA+++ +MK+CI AQD+C+R IQLAGYDVYY
 Sbjct: 75 IYETGIRIPTICFSGHRRYPLGSNDPAIEAKSLKLMKQCIELAQDLGVRTIQLAGYDVYY 134

Query: 121 EEKSPETRARFIKNLRQACTWAEAAQVILSIEIMDDPFMNSIEKYLAKEIDSPYLFVY 180
 E+KSPETRARFIKNLRQ+C WAEAAQV+LSIEIMDDPF+NSIEKYLAKEIDSPYLFVY
 Sbjct: 135 EKKSPETRARFIKNLRQSCDWAEAAQVMSLSIEIMDDPFINSIEKYLAKEIDSPYLFVY 194

Query: 181 PDTGNVSAWHNDLWSEFYNGHRSIAALHIKDTYAVTETSKGQFRDVPFGQGCVDWEEMFA 240
 PD GNVSAWHNDLWSEFYNGH+SIAALH+KDTYAVTETSKGQFRDVPFGQGCVDW+E+FA
 Sbjct: 195 PDAGNVSAWHNDLWSEFYNGHKSIAALHLKDTYAVTETSKGQFRDVPFGQGCVDWQELFA 254

Query: 241 VIKKTNYNGPFLIEMWSENCETVEETRAAIKEAQDFLYPLMEKTGV 286
 V+KKTNYNGPFLIEMWSENC+TVEET+AAIKEAQDFLYPL+EK G+
 Sbjct: 255 VLKKTNYNGPFLIEMWSENCETVEETKAAIKEAQDFLYPLIEKAGL 300

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1493

A DNA sequence (GBSx1579) was identified in *S.galactiae* <SEQ ID 4597> which encodes the amino acid sequence <SEQ ID 4598>. This protein is predicted to be L-ribulose 5-phosphate 4-epimerase. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

-1651-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2559(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD45716 GB:AF160811 L-ribulose 5-phosphate 4-epimerase
 [Bacillus stearothermophilus]

Identities = 143/229 (62%), Positives = 176/229 (76%), Gaps = 2/229 (0%)

Query: 5 LQEMRERVCEANKSLPVSLSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMVVTDL 64
 L+E+++ V EAN LP + LV FTWGNVS +DRE GL+VIKPSGV YD+LT ++MVV DL
 Sbjct: 2 LLELKQAVLEANLQLPQYRLVFTWGNVSGIDRERGLVVIKPSGVAYDKLTIDDMVVVDL 61

Query: 65 EGNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHSTHSTEAVGWAQAGRDIPFYGTTHADYF 124
 GN+VEGDL PSSD PTH+ LYK +P +GGIVHSTHST A WAQAG+ IP GTTHADYF
 Sbjct: 62 TGNVVEGDLKPSSDTPTHVLYKQFPGIGGIVHSTHSTWATVWAQAGKGPALGTTHADYF 121

Query: 125 YGPVPCARSLSEDEVNTAYEKETGSGVIIIEEFERRDLDPMAVPGIVVRNHGPFTWGDPAQ 184
 YG +PC R ++ +E+ AYE ETG VI E F R LDP+ +PG++V HGPFT WGDPA
 Sbjct: 122 YGEIPCTRPMTNEEIQGA YELETKGVITETF--RFLDPLQMGVLVHGHGPFPAWGKDPAN 179

Query: 185 AVYHVSVVLEEVAKMNRFTQINPRVEPAPKYIMDKHYLRKHGPNAYYGQ 233
 AV+++VVLEEVAKM T +NP +P + ++D+HYLRKHG NAYYGQ

Sbjct: 180 AVHNAVSVLEEVAKMAARTYMLNPNAPKISQTLDRHYLRKHGANAYYGQ 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4599> which encodes the amino acid sequence <SEQ ID 4600>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2257(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 207/234 (88%), Positives = 220/234 (93%)

Query: 1 MAKSLQEMRERVCEANKSLPVSLSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMV 60
 MAK+LQEMRERVC ANKSLP H LVKFTWGNVSEV RE G IVIKPSGVDYD LTPENMV
 Sbjct: 1 MAKNLQEMRERVCAANKSLPQHGLVKFTWGNVSEVCRELGRIVIKPSGVDYDLLTPENMV 60

Query: 61 VTDLEGNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHSTHSTEAVGWAQAGRDIPFYGTTH 120
 VTDL+GN+VEGDLNPSSDLPTHV+LYKAWPEVGGIVHSTHSTEAVGWAQAGRDIPFYGTTH
 Sbjct: 61 VTDLDGNVVEGDLNPSSDLPTHVELYKAWPEVGGIVHSTHSTEAVGWAQAGRDIPFYGTTH 120

Query: 121 ADYFYGPVPCARSLSEDEVNTAYEKETGSGVIIIEEFERRDLDPMAVPGIVVRNHGPFTWKG 180
 ADYFYGPVPCARSL++ EV+ AYE+ETG+VI+EEF +R LDPMAVPGIVVRNHGPFTWKG
 Sbjct: 121 ADYFYGPVPCARSLTKAEVDGAYEQETCNVILEEFKRGDLDPMAVPGIVVRNHGPFTWKG 180

Query: 181 DPAQAVYHVSVVLEEVAKMNRFTQINPRVEPAPKYIMDKHYLRKHGPNAYYGQK 234
 P QAVYHVSVVLEEVA+MNR TEQINPRVEPAP+YIMDKHYLRKHGPNAYYGQK
 Sbjct: 181 TPEQAVYHVSVVLEEVARMNRLTEQINPRVEPAPRYIMDKHYLRKHGPNAYYGQK 234

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1652-

Example 1494

A DNA sequence (GBSx1580) was identified in *S.galactiae* <SEQ ID 4601> which encodes the amino acid sequence <SEQ ID 4602>. This protein is predicted to be transaldolase (tal). Analysis of this protein sequence reveals the following:

5 Possible site: 45
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.4232(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10149> which encodes amino acid sequence <SEQ ID 10150> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

 >GP:AAB98962 GB:U67539 transaldolase [Methanococcus jannaschii]
 Identities = 124/214 (57%), Positives = 157/214 (72%)

20 Query: 19 MKYFLDTADVSEIRRLNRLGIVDGVTTNPTIISREGRDFKEVINEICQIVDGPVSAEVTG 78
 MK+FLDTA+V EI++ LG+VDGVTTNPT++++EGRDF EV+ EIC+IV+GPVSAEV
 Sbjct: 1 MKFFLDTANVEEIKKYAELGLVDGVTTNPTLVAKEGRDFYEVVKEICEIVEGPVSAEVIS 60

 Query: 79 LTCDEMVTEAREIAKWSPNVVKIPMTTEGLAAVSQLSKEGIKTNVTILIFTVAQGLSAMK 138
 + MV EARE+AK + N+V+KIPMT++G+ AV LS EGIKTNVTIL+F+ Q L A K
 25 Sbjct: 61 TDAEGMVKEARELAKLADNIVIKIPMTKDGMAKVKILSAEGIKTNVTILVFSPLQALVAAK 120

 Query: 139 AGATFISPFVGRLEDIGTDAYALIRDLRHIIDFYGFQSEIIAASIRGLAHVEGVAKCGAH 198
 AGAT++SPFVGRL+DIG LI D+ I Y ++E+I AS+R HV AK GA
 30 Sbjct: 121 AGATYVSPFVGRLDDTGHVGMKLIEDVVKIYKNYDIKTEVIVASVRHPWHVLEAAKIGAD 180

 Query: 199 IATIPDKTFASLFTHPLTDKGIETFLKDWDSFKK 232
 IAT+P LF HPLTD G+E FLKDWD + K
 35 Sbjct: 181 IATMPPAVMDKLEFNHPLTDIGLERFLKDWDEYLK 214

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4603> which encodes the amino acid sequence <SEQ ID 4604>. Analysis of this protein sequence reveals the following:

 Possible site: 15
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1902(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below.

 Identities = 162/214 (75%), Positives = 180/214 (83%)

50 Query: 19 MKYFLDTADVSEIRRLNRLGIVDGVTTNPTIISREGRDFKEVINEICQIVDGPVSAEVTG 78
 MK+FLDTA+V+ I+ +N LG+VDGVTTNPTIISREGRDF+ VI EIC IVDGP+SAEVTG
 Sbjct: 1 MKFFLDTANVAAIKAINELGVVDGVTTNPTIISREGRDFETVIKEICDIVDGPISAEVTG 60

 Query: 79 LTCDEMVTEAREIAKWSPNVVKIPMTTEGLAAVSQLSKEGIKTNVTILIFTVAQGLSAMK 138
 LT D MV EAR IAKW NVVVKIPMT EGL A + LSKEGIKTNVTILIFTV+QGL AMK
 55 Sbjct: 61 LTADAMVEEARSIAKWHDNVVVKIPMTTEGLKATNLSKEGIKTNVTILIFTVSQGLMAMK 120

 Query: 139 AGATFISPFVGRLEDIGTDAYALIRDLRHIIDFYGFQSEIIAASIRGLAHVEGVAKCGAH 198
 AGAT+ISPF+GRLEDIGTDAY LI DLR IID Y FQ+EIIAASIR AHVE VAK GAH
 60 Sbjct: 121 AGATYISPFVGRLEDIGTDAYQLISDLREIIDLYDFQAEIIAASIRTTAHVEAVAKLGAH 180

-1653-

Query: 199 IATIPDKTFASLFTHTPLTDKGIETFLKDWDSEFKK 232
 IATIPD FA + HPLT G++TF++DW SFKK
 Sbjct: 181 IATIPDPLFAKMTQHPLTTNGLKTPMEDWASFKK 214

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1495

A DNA sequence (GBSx1581) was identified in *Sagalactiae* <SEQ ID 4605> which encodes the amino acid sequence <SEQ ID 4606>. Analysis of this protein sequence reveals the following:

10 Possible site: 22
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1263(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14129 GB:Z99115 transcriptional regulator (LacI family)
 [Bacillus subtilis]
 Identities = 108/331 (32%), Positives = 188/331 (56%), Gaps = 12/331 (3%)

20 Query: 6 TISDIANLVGVSKATVSYLNGNYKKMSLQTKKIRLAIKETGYQPSKIAQSLVTKNTRT 65
 TI D+A GVS+K+TVS Y+NG +S + + I+ AI E Y+PSK+AQ L K ++
 25 Sbjct: 10 TIKDVAECAGVSKSTVSRYINGKIDAISEPKVKNIKKAI AELNYRPSKMAQGLKIKKSKL 69

Query: 66 IGVVIADITNPFISSVMKGIHDTCCQFGYSVNFTNSDNDIDIELENLNRLNQNVSGIIL 125
 IG V+ADITNPF + +G+ + C Q+GYS+ N+DN + E E L +L +V G+IL
 30 Sbjct: 70 IGFVVADITNPFSSVAAPRGVEEVCDQYGSIMVCNTDNSPEKEREMLLKLEAHSVEGLIL 129

Query: 126 DSVDPNHSFIETLSNDRL--VMVDQAKDIKVDTVASDNKESTQIFLEKMQEAGYHDIYF 183
 ++ N + + ++ +++DR+ D+K+DTV +DN+ T+ L+K+ GY D+
 35 Sbjct: 130 NATGENKDVLRFAEQQIPTILIDRKLPLKLDVTVDNRWITKEILQKVYSGYTDVAL 189

Query: 184 VTYPGISTRELRYEGFKEVVS-SNPDKLIIITE-DGSTQRILDI-----IEHSEQKP 235
 T PI IS R R ++E+ S N + L+ + E D + L E EQK
 40 Sbjct: 190 FTEPISSISPRAEAAVYQEMASVQNVNGLVRLHEIDVKDKQLKAE LRSFHKEMPEQKK 249

Query: 236 GFLMNGPTLLNFMKLNQSTVSYPEDYGLGSYEDLEWMQVLTNPVSCIKQDSYGIGCLA 295
 L +NG +L + + + + P+D G+ ++D EW +++ P ++ I Q S+ +G A
 45 Sbjct: 250 AILALNGLIMLKII SCMBELGLRIPQDIGIAGFDDTEWYKLI GPGITTIAQPSHDMGRTA 309

Query: 296 AQCLIEKISQNEPTTARLLEVKNIQIVIRQS 326
 + ++++I + + +E++ +++R+S
 50 Sbjct: 310 MERV LKRIE--GDKGAPQTIELEAKVIMRKS 338

There is also homology to SEQ ID 2366.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

50 Example 1496

A DNA sequence (GBSx1582) was identified in *Sagalactiae* <SEQ ID 4607> which encodes the amino acid sequence <SEQ ID 4608>. Analysis of this protein sequence reveals the following:

55 Possible site: 40
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1654-

bacterial cytoplasm --- Certainty=0.1661 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1497

10 A DNA sequence (GBSx1583) was identified in *S.agalactiae* <SEQ ID 4609> which encodes the amino acid sequence <SEQ ID 4610>. This protein is predicted to be GLYCERATE DEHYDROGENASE. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

15

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB50351 GB:AJ248287 GLYCERATE DEHYDROGENASE [Pyrococcus abyssi]
 Identities = 123/325 (37%), Positives = 192/325 (58%), Gaps = 8/325 (2%)

25

Query: 1 MDKKILVTGIVPKEGLRKLMDRFDVTYSED-RPFSRDYVLEHLSEYDGLWLLM-GQKGDK 58
 M K ++ +T +P+ G+ L F+V ED R R+ +LE ++ +D + M ++ D+
 Sbjct: 1 MSKPRVFITREIPEVGIEMLEKEFEVEVWEDEREIPREILLEKVKDQVDALVTMLSERIDR 60

30

Query: 59 EMIDAGENLQIISLNAVGFHDVDTAYAKEKGIIVSNPQAVRVPTAEMTFALILAASKRL 118
 E+ + L+I++ AVG+D++D A ++GI V+N+P + TA++ FAL+LA ++ L
 Sbjct: 61 EVFERAPRLRIVANYAVGYDNIDVEEATKRGIVYVINTPGVLTADLAFALLATARHL 120

35

Query: 119 AFYDSIVRSGEW----IDPSEQRYQGLTLQGSTLGIYGMGRIGLTVANFAKAFGMTVVYN 174
 D RSGEW + + + G + G T+GI G GRIG +A A+ F M ++Y
 Sbjct: 121 VKGDKPTRSGEWKKRGVAVHPKWFLGYDVYKTIIGIIGFGRIGQAIKRARGFDMRILYY 180

40

Query: 175 DVYRLPEDKEKELGVTYLEFDQLIKTADVITIHPALPSTIHKFNKDVFAKMKNRSYLIN 234
 R PE EKEL + D+L++ +D + + P T H N++ MK + LIN
 Sbjct: 181 SRTKPE-VEKELNAEFKPLDELLRESDFVVLAVPLNKETYHMINERLKMMLTAILIN 239

45

Query: 235 AARGPIVSEALIEALKEGEIAGAGLDVFNENPQVSEGLRSLDNVIMSPHAGTGITIEGRR 294
 ARG ++ +ALI+ALKEG IAGAGLDV+E EP +E L SLDNV+++PH G+ T R
 Sbjct: 240 VARGKVIDTKALIKALKEGWIAGAGLDVYEEEPYNEELFSLDNVVLTPHIGSATFGARE 299

Query: 295 TLAEAAADNIIAFFDGK-PQIVNKK 318
 +A+ A+N+IAF G+ P +VN+
 Sbjct: 300 GMAKLVAENLIAFKRGEVPPTLVNR 324

There is also homology to SEQ ID 124.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1655-

Example 1498

A DNA sequence (GBSx1585) was identified in *S.agalactiae* <SEQ ID 4611> which encodes the amino acid sequence <SEQ ID 4612>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.1898(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1499

A DNA sequence (GBSx1586) was identified in *S.agalactiae* <SEQ ID 4613> which encodes the amino acid sequence <SEQ ID 4614>. This protein is predicted to be PTS system, galactitol specific IIC component. Analysis of this protein sequence reveals the following:

```

20  Possible site: 25
   >>> Seems to have no N-terminal signal sequence

   INTEGRAL Likelihood = -13.27 Transmembrane 254 - 270 ( 245 - 277)
   INTEGRAL Likelihood = -9.24 Transmembrane 77 - 93 ( 71 - 100)
25  INTEGRAL Likelihood = -9.24 Transmembrane 367 - 383 ( 364 - 386)
   INTEGRAL Likelihood = -8.28 Transmembrane 32 - 48 ( 26 - 54)
   INTEGRAL Likelihood = -7.38 Transmembrane 186 - 202 ( 182 - 215)
   INTEGRAL Likelihood = -6.26 Transmembrane 158 - 174 ( 151 - 180)
   INTEGRAL Likelihood = -5.79 Transmembrane 279 - 295 ( 276 - 296)
30  INTEGRAL Likelihood = -1.12 Transmembrane 342 - 358 ( 342 - 359)
   INTEGRAL Likelihood = -0.00 Transmembrane 308 - 324 ( 308 - 324)

   ----- Final Results -----
   bacterial membrane --- Certainty=0.6307(Affirmative) < succ>
35  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8825> which encodes amino acid sequence <SEQ ID 8826> was also identified. Analysis of this protein sequence reveals the following:

```

40  Lipop: Possible site: -1 Crend: 9
   McG: Discrim Score: 8.30
   GvH: Signal Score (-7.5): 2.97
   Possible site: 58
   >>> Seems to have a cleavable N-term signal seq.
   ALOM program count: 9 value: -13.27 threshold: 0.0
45  INTEGRAL Likelihood = -13.27 Transmembrane 321 - 337 ( 312 - 344)
   INTEGRAL Likelihood = -9.24 Transmembrane 144 - 160 ( 138 - 167)
   INTEGRAL Likelihood = -9.24 Transmembrane 434 - 450 ( 431 - 453)
   INTEGRAL Likelihood = -8.28 Transmembrane 99 - 115 ( 93 - 121)
   INTEGRAL Likelihood = -7.38 Transmembrane 253 - 269 ( 249 - 282)
50  INTEGRAL Likelihood = -6.26 Transmembrane 225 - 241 ( 218 - 247)
   INTEGRAL Likelihood = -5.79 Transmembrane 346 - 362 ( 343 - 363)
   INTEGRAL Likelihood = -1.12 Transmembrane 409 - 425 ( 409 - 426)
   INTEGRAL Likelihood = -0.00 Transmembrane 375 - 391 ( 375 - 391)
55  PERIPHERAL Likelihood = 0.69 188
   modified ALOM score: 3.15

```

-1656-

*** Reasoning Step: 3

----- Final Results -----

5 bacterial membrane --- Certainty=0.6307(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP: BAB03909 GB: AP001507 PTS system, galactitol-specific enzyme II,
 C component [Bacillus halodurans]
 Identities = 92/347 (26%), Positives = 173/347 (49%), Gaps = 15/347 (4%)

15 Query: 1 MVKTTGLHLPIVDIGWQAGSLTAFSSEIGLSFFVFGLLIELGLFLLGITRVFVPSNLWNN 60
 MV G+ L ++D+GW A S A++S + GL++ + + + T+ + ++WN
 Sbjct: 70 MVDRLGVDLNVIDVGWPATSSIAWASVVAAFIPLGLIVNVIMLVTKTKT-MNVDIWNF 128

20 Query: 61 FGYYMIGTMAYAAATGNFILSFAPMVFVLLYSVMSEVLADRWSEYYGVKNATINSIHNE 120
 + Y + Y + + I + V + +L+++ A SE+Y + +I + I
 Sbjct: 129 WHYTFMAAVVYTVSDSIIQALIAAVMFQIVALKVADWTAPMVSEFYELPGVSIATGSTIS 188

25 Query: 121 TLIPALILDPLWNLLGVNKKVKNPESLKTGLGIFGEPMTLGFILGVIIGVLGSLRNLASI 180
 ++ + + G+ +P++++ + GIFGE + +G ILG IG+L
 Sbjct: 189 YAPGIWLKGIQKIPGIKHNADPDTIQRRFGIFGESIFIGLILGAAIGLLAGYNV---- 244

30 Query: 181 DTWGGILGFAVALAAMVTIFPLITGVFASAFAPLAEAVERNKKKESQAEQGALDKKRWFI 240
 G ++ +A+AAVM + P + + P++E+ K + I
 Sbjct: 245 ---GEVIEIGMAAAVMVLMPRMKILMEGLMPVSESAREWLNKR-----FGDREIHI 294

35 Query: 241 AVDDGVGFGEPATIIAGLILVPMVVISLILPGNEALPVVDLIAIPFMIEAMIAVSKGNI 300
 +D V G P+ I LILVP+ V++++ILPGN LP DL IPF++ ++ ++GNI
 Sbjct: 295 GLDAAVLLGHPSVISTALILVPLTVLLAVILPGNALLPFGDLATIPFIVAFIVGAARGNI 354

Query: 301 LKAILNGIIFWFSGLYASALGPIYTEAVKHFGTALPAGVTLMISFN 347
 + ++L G I +L LY A+ + P++T+ ++ +P G LI S +
 Sbjct: 355 IHSVLAGAIMIALSLYMATDIAPVFTKMAENSNNFMPEGSALISSID 401

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1500

40 A DNA sequence (GBSx1587) was identified in *S.agalactiae* <SEQ ID 4615> which encodes the amino acid sequence <SEQ ID 4616>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1013(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1657-

Example 1501

A DNA sequence (GBSx1588) was identified in *S.agalactiae* <SEQ ID 4617> which encodes the amino acid sequence <SEQ ID 4618>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1294(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10147> which encodes amino acid sequence <SEQ ID 10148> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC76604 GB:AE000435 L-xylulose kinase, cryptic [Escherichia coli K12]

Identities = 156/496 (31%), Positives = 261/496 (52%), Gaps = 18/496 (3%)

Query: 16 YYLSIDYGGTNTKALIFDKLGHQIAVSSFETLKNETQSGHRQVNLVKTWNAITSAIREVI 75

Y+L +D GG+ KA ++D+ G + V Q G + ++ + W + IR ++

Sbjct: 4 YWLGLDGGSWLKAGLYDREGREAGVQRLPLCALSPQPGWAERDMAELWQCCMAVIRALL 63

Query: 76 QISKLSPEQISAVACIGHGKGLYLLDNKLEPLEQGILSTDNRADLAQYFESK--LDNIW 133

S +S EQI + GKGL+LLD +PL ILS+D RA ++ + ++ + ++

Sbjct: 64 THSGVSGEQIVGIGISAQGGKGLFLLDKNDKPLGNAILSSDRAMEIVRRWQEDGIPEKLY 123

Query: 134 ELTRQHIFPSQSPVILRWLKDYPETYKSIGAVLSAKDFIRYKLTGKVQQEYGDASGNHW 193

LTRQ ++ +LRWLK+++PE Y IG V+ D++R+ LTG E + S ++

Sbjct: 124 PLTRQTLTWGHPVSLLRWLKEHEPERYAQIGCVMMTHDYLRWCLTGKGCESNISSENL 183

Query: 194 INFQTGTYPAILDFFGIREIENSLPELIDSADLVPGGISSQAAKETGLVEGTVPVVGGLF 253

N G YDP + D+ GI EI ++LP ++ SA++ G I++Q A TGL GTPVVGGLF

Sbjct: 184 YNMSLGEYDPCLTDLNGLAEINHALPPVVGSAIC-GEITAQTAALTGLKAGTVPVVGGLF 242

Query: 254 DIDACALGSGVLESDFSVISGTWNINT--YPSLKPAKQDSGLMTSYFPDRRYLLEASSP 311

D+ + AL +G+ + T + + GTW + + L+ + + Y D +++ +SP

Sbjct: 243 DVVSTALCAGIEDEFTLNAVMTWAVTSGITRGLRDGEAHPYVYGRYVNDGEFIVHEASP 302

Query: 312 TSAGNINFMKMLMHQEIDNAKSSGGSIDNLEEFLLTHDATHHGLIFFPFLYGSNTSQD 371

TS+GNL + G +D + + + L F PFLYGSN +

Sbjct: 303 TSSGNLEWF-----TAQWGEISFDEINQAVASLPKAGGDLFFLPFLYGSNAGLE 351

Query: 372 ASACFFGLTTKSTKSQMIRAVYEGIAFAHKQHITDLIKSRGSVPKIIRFSGGATNSPAWM 431

++ F+G+ T++ +++A+YEG+ F+H H+ + ++ R + +R +GG +S WM

Sbjct: 352 MTSGFYGMQAIHTRAHLQAIYEGVVFSMTHL-NRMRRERFTDVHTLRVTGGPAHSDVWM 410

Query: 432 QMFSDILNFPPIETVEGTELGGGLGAILARHALDKI-SLKEAVQDMVRVKAIYKPQLSEVK 490

QM +D+ IE + E G G A+ AR + EA +D+ P ++ +

Sbjct: 411 QMLADVSGLRIELPQVEETGCFGAALAAARVGTGVYHNFSEAQRDLRHPVRTLLPDMTAHQ 470

Query: 491 GYKKKYHAYQKILETL 506

Y+KKY YQ L+ L

Sbjct: 471 LYQKKYQRYQHLLAAL 486

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1658-

Example 1502

A DNA sequence (GBSx1589) was identified in *S.agalactiae* <SEQ ID 4619> which encodes the amino acid sequence <SEQ ID 4620>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG05648 GB:AE004652 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 59/235 (25%), Positives = 104/235 (44%), Gaps = 9/235 (3%)

Query: 23 QVQLIKLVKDLGFSRFEIRQELLQDDPRELPALKAEADFYDINLYSANEDLIK-GGKVN 81
 Q + L+ G R E+R+EL P + AL A + +S+ +L+ G++N
 Sbjct: 23 QASFLPLLAMAGAQRVLEELFAGPP-DTEALTAAIQLQGLECVFSSPLELWREDGQLN 81

Query: 82 PYLNKGLKEASQLGAPFIKLVNGQTRNLSKEELEPLKEILKSQTIGIKVENNQDPKAATV 141
 P L L+ A GA ++K+++G + +L L L + + VEN+Q P+ +
 Sbjct: 82 PELEPTLRRAEACGAGWLKVSLLGLPE--QPDLAALGRRIARHGLQLLVENDQTPQGGRI 139

Query: 142 ENCQYFMTLVKELQIPISFVFDTANWAFINQDLYQAVNMLACDTTYLHCKNFIQVAGKPH 201
 E + F L + Q+ ++ FD NW + Q +A L Y+HCK I+
 Sbjct: 140 EVLERFFRLAERQQDLAMTFDIGNWRWQEQAADAAALRLGRYVGVYHCKAVIRNRDGKL 199

Query: 202 LSKSLFEGEINLTD-LLKSFSNCEYLALPYTE----LEILKRDVQRLISISNSQ 251
 ++ ++ LL+ F A+EYP + L + +R + L + Q
 Sbjct: 200 VAVPPSAADLQYWQRLQLHFPEGVARAIEYPLQGGDLLSLSRRHIAALARLGQPQ 254

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1503

A DNA sequence (GBSx1590) was identified in *S.agalactiae* <SEQ ID 4621> which encodes the amino acid sequence <SEQ ID 4622>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0430 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB03939 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 136/511 (26%), Positives = 234/511 (45%), Gaps = 29/511 (5%)

Query: 4 LDKKSYDLLFYLLKLEPETVMAIANALNQSRKVVYHLEKINDALPSDVPQIVSYPRV- 62
 LD++S +L LL + + LN SRR VY LEKIN L + V R
 Sbjct: 3 LDQRSTFILTQLLHARSYLP IQELTQKLVNSRRRTVYNDLEKINSWLEEQGLKAVYKVR SQ 62

Query: 63 GILLTEKQKAACRLLLDDEVTDYSYVMKSSERLQSLVSIIVAKDRVTIDRLMQLNDVSRN 122
 G++L E+ K L + + Y + ER ++ ++ + + ++ LM VSRN
 Sbjct: 63 GLILDERAKEEIPTKRLSLKSWHYEYSAQERKAWVVIYLLTRLEPLFLEHLMDRTCVSRN 122

-1659-

Query: 123 TILNDLNELRSELAKEYNLQLOSTKCRGYFLDGHPL----SIIQYLYKLLDDIYHNGSS 178
 T ++D+ L+ EL ++L L+ + GY + G +++ YL + L
 Sbjct: 123 TTIDDIKCLKDEL--NNFHLALEFERKIDGYTISGDETDKRRKALVYVLSQALPQQNWETEL 180

5 Query: 179 SFIDLFNHKLSQAFGASTYFSKEVLDYFHHYLFISQSLGKKINSQDQGMILPFI 238
 S I +F L F + E L + + S++ L KI D L F+L
 Sbjct: 181 SPIRIF---LRTKRDNGRIFTIEELQKVYDVISESEKVL--KIQTDDVLHSLSLRFLLF 235

10 Query: 239 AYRK-----MRLSPEVQTSLSNDFSLVWQKEYEIAKELADELEENFQLSLDEIEVGLVA 293
 R +++ P + L KEYE AK ++ +LE+ F + + EV +
 Sbjct: 236 MKRVAKGKFIKVHPLEKQVLKGT-----KEYEAAKVMSPKLEQAFGVHYPDEEVLYLT 288

15 Query: 294 MLMLSPFRKDRDN-HLESQ-DYDDMRATLTSFLKELEERYHLHFVHKDILLRQLLTHCKAL 351
 +LS + + N +ES+ + ++ +TS + + ++ + F K+ L + L H K
 Sbjct: 289 THILSSKINYANGEIESRKESQELTHIVTSMVNDFOKYACVVFEKELLEKNLFFHIKPA 348

20 Query: 352 LYRKRYGIFSVNPLTEHIKDKYEELFAITSSSVKLEKAWQIKLTDVAYLTIHLGGEL 411
 YR +YG+ N + E IK Y ELF +T V LE+ + D++VA++T+H G +
 Sbjct: 349 FYRIKYGLEVENNIAESIKTSYPELEFLLRKVVHYLERYVGKSVNDNEVAFITMHFVGM 408

25 Query: 412 RNSQQSPNK-LKLIVISDEGIAIQKLLKQCQRYLTNSDIEAVFTTEQYQSVSDLMHVD 470
 R P K K +IV G+ + L Q + DI + +Y+ + VD
 Sbjct: 409 RREGTIPTKRKKALIVCANGVGTSQLKQLEGLFPAVDIITCSIREVEKTP--VEVDF 466

Query: 471 VVSTSDALESRFPMLVVHPVITDDDIIRLIR 501
 ++ST+ E P+ +V+P+LT+ + RL++
 Sbjct: 467 IISTTSIPEKNVPFIVNPILTETEKERLLK 497

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4623> which encodes the amino acid
 30 sequence <SEQ ID 4624>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.0745(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

40 Identities = 368/548 (67%), Positives = 456/548 (83%)

Query: 1 MIILDKKSYDLLFYLLKLEPETVMAIANALNQSRKRVYHLEKINDALPSDVPQIVSY 60
 M+ILDKKSYDLL YLLKLE PETVMAI++ALNQSRKRVY L+KIN ALP V QI+SY
 Sbjct: 1 MMILDKKSYDLLSYLLKLETPETVMAISHALNQSRKRVYQLDKINQALPKGVDQIISY 60

45 Query: 61 RVGILLTEKQKAACRLLLEEDVTDYSYVMKSSERLQSLVSIVVAKDRVTIDRLMQINDVS 120
 R+GILLT OKAACRLLL+EVTDY+YVMKS ER +LS + I V+ +RVTID+LMQ+NDVS
 Sbjct: 61 RLGILLTADQKAACRLLLEEDVTDYNYVMKSDERRRLSSIYIAVSTERTVIDKLMQINDVS 120

50 Query: 121 RNTILNDLNELRSELAKEYNLQLOSTKCRGYFLDGHPLSIIQYLYKLLDDIYHNGSSSF 180
 RNTILNDL ELR EL +K+Y +QL +TK RGY+ HP+++IQYLYKLL D+Y G++SF
 Sbjct: 121 RNTILNDLTELRELEDKQYKIQLHATKARGYYFGCHPMALIQYLYKLLVDVYGGNTSF 180

55 Query: 181 IDLFNHKLSQAFGASTYFSKEVLDYFHHYLFISQSLGKKINSQDQGMILPFI 240
 ID+FN KLS+ G S YFSK++L YFH YLF+SQ SLGK IN+QD QFM+QILPF+L++Y
 Sbjct: 181 IDIFNRKLSEIQGLSVYFSKIDILTYFHEYLFSLQASLGKTINTQDSQFMLQILPFMLLSY 240

60 Query: 241 RKMRLSPEVQTSLSNDFSLVWQKEYEIAKELADELEENFQLSLDEIEVGLVAMLMLSFR 300
 R MRL E +++L +F L+W+RKEY IA++LA EL NF+L LD+IEV +VAMLMLSFR
 Sbjct: 241 RNMRLDSETKSALKQEPHLIWRKEYHIAQDLARELYHNFKLHDDIEVSMVAMLMLSFR 300

65 Query: 301 KDRDNHLESQDYDDMRATLTSFLKELEERYHLHFVHKDILLRQLLTHCKALLYRKRYGIF 360
 KD+D+H+ESQDYDDMRAT++ F+ +LE RY LHF HK+DLL++L THCKAL+YRK YGIF
 Sbjct: 301 KDQDHVESQDYDDMRATISHFIDQLESRYQLHFTHKQDLLKRLTHCKALVYRKAYGIF 360

-1660-

Query: 361 SVNPLTEHIKDKYEELFAITSSSVKLEKAWQIKLTDDDVAYLTIHLGGELRNSQQSPNK 420
 VNPLT+H+K+KYEELFA+T S +LE+ W I LTDDD+AYLTIHLGGELR++ K
 Sbjct: 361 LVNPLTDHVKEKYEELFAMTQSCATILEQDWTISLTDDDIAYLTIHLGGELRHNNTQEK 420

5 Query: 421 LKLVIVSDEGIAIQKLLKQCQRYLTNSDIEAVFTTEQYQSVSDLMHVDMMVSTSDALES 480
 KLVIVSD+GI IQKLL KQCQRYL N IEAVFTTEQYQSV DL+ VDM+V+T+D L++
 Sbjct: 421 TKLVIVSDDGIGIQKLLFKQCQRYLANGQIEAVFTTEQYQSVYDLLAVDMIVATDTLKT 480

10 Query: 481 RFPMLVVHPVLTDDDIIRLIRFSKKGNCANSNQFTNELEKTIAQYVKEDSERYVLKSKIE 540
 + PML+V+P+L+DDDI+LIRFSK+G + ++F+ EL K I VK++S+RY L SKIE
 Sbjct: 481 KIPMLIVNPILSDDDIIRLIRFSKQGRLEHRSRSTELTKAIEAVVKDESRYALVSKIE 540

Query: 541 KLIHQELL 548
 KLIH+ELL
 15 Sbjct: 541 KLIHRELL 548

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1504

20 A DNA sequence (GBSx1591) was identified in *Sagalactiae* <SEQ ID 4625> which encodes the amino acid sequence <SEQ ID 4626>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2692(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77149 GB:AE000491 orf, hypothetical protein [Escherichia coli K12]
 Identities = 211/363 (58%), Positives = 270/363 (74%), Gaps = 9/363 (2%)

35 Query: 1 MPNVKDITRESWILSTFPPEWGTWLNNEIEEEVVAEGNFAMWWLGNCGVWIKTPGGANVVM 60
 M VK ITRESWILSTFPPEWG+WLNEEIE+E VA G FAMWWLG G+W+K+ GG NV +
 Sbjct: 3 MSKVKSITRESWILSTFPPEWGSWLNNEIEEQEQVAPGTFAMWWLGCTGIWLKSEGNTNVCV 62

40 Query: 61 DLWSNRGKSTKKVKDMVRGHQMANMAGVRKLQPNLRAQPMVIDPFAINELDYLVSHFHS 120
 D W GK + M +GHQM MAGV+KLQPNLR P V+DPFAI ++D L +H H+
 Sbjct: 63 DFWCGTGKQSHGNPLMKQGHQMORMAGVKKLQPNLRTTPFVLDPPFAIRQIDAVLATHDHN 122

45 Query: 121 DHIDINTAAAIINNPNLDHVKFVGPYECGHIWKKWGVPEERIIIVIKPGESFEFKDIKVT 180
 DHID+N AAA++ N D V F+GP C ++W WGVP+ER IV+KPG+ + KDI++ A
 Sbjct: 123 DHIDVNVAAMQNC-ADDVPFIGPKTCVDLWIGWGVPKERCIVVKPGDVVKVDIEIHA 181

50 Query: 181 VESFDRTCLVTLFVDGAEEDGELAGLAVTDEEMARKAVNYIFETPGGTTIYGADSHFSN 240
 +++FDRT L+TLP D + AG V + M +AVNY+F+TPGG++YH DSH+SN
 Sbjct: 182 LDAFDRTALITLPADQ-----KAAG--VLPDGMDDRAVNYLFKTPGGSLYHSGDSHYSN 233

55 Query: 241 YFAKHGKDYKIDVAINNYGDNPNVGIQDKMTSIDLLRMAENLRAKVIIIPVHYDIWSNFMAS 300
 Y+AKHG +++IDVA+ +YG+NP GI DKMTS D+LRM E L AKV+IP H+DIWSNF A
 Sbjct: 234 YYAKHGNEHQIDVALGSYGENPRGITDKMTSADMLRMGEALNAKVVIIPFHDIWSNFQAD 293

Query: 301 TDEILQLWKMKEKRLQYDFHPFIWEVGGKYTPQDKDRIEYHHPRGFDDCFEQESNIQFK 360
 EI LW+M+K+RL+Y F PFIW+VGGK+T+P DKD EYH+PRGFDDCF E ++ FK
 Sbjct: 294 PQIRVLWEMKKDRLKYGKPFPIWQVGGKFTWPLDKDNFEYHYPRGFDDCFETIEPDLFPK 353

Query: 361 ALL 363
 + L
 60 Sbjct: 354 SFL 356

-1661-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4627> which encodes the amino acid sequence <SEQ ID 4628>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3298(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 315/363 (86%), Positives = 348/363 (95%)

Query: 1 MPNVKDITRESWILSTFPEWGTWLNNEEIEEEVVAEGNFAMWWLGNCGVWIKTPGGANVVM 60

M V+DITRESWIL+TFPEWGTWLNNEEIE+EVV NFAMWWLGNCG+WIKTPGGANVVM

Sbjct: 1 MTKVQDITRESWILNTFPEWGTWLNNEEIEQEVVPADNFAMWWLGNCGIWIKTPGGANVVM 60

Query: 61 DLWSNRGKSTKKVKDMVRGHQMANMAGVRKLQPNLRAQPMVIDPFAINELDYLVSHFHS 120

DLWSNRGK+TK+VKDMVRGHQMANMAG RKLQPNLRAQPMVIDPF INELDYLVSH+HS

Sbjct: 61 DLWSNRGKATKQVKDMVRGHQMANMAGARKLQPNLRAQPMVIDPFMINELDYLVSHYHS 120

Query: 121 DHIDINTAAAIINNPNLDHVKFVGPYECGEIWKKGWVPEERIIIVIKPGESFEFKDIKITA 180

DHIDINTAAAIINNPN L+HVKFVGPYECGE+WK WGVV++RI+++KPG+SFEFKDIK+TA

Sbjct: 121 DHIDINTAAAIINNEKLNHVKFVGPYECGEVWKNWGVKDRIMILKPGDSFEFKDIKITA 180

Query: 181 VESFDRITCLVTLPLVDGAEHDEGLAVTDEEMARKAVNYIFETPGGTIYHGADSHFSN 240

VESFDRITCLVTLPL+ GA+ DG+LAGLA+TD++MARKAVNYIFETPGGTIYHGADSHFSN

Sbjct: 181 VESFDRITCLVTLPLIQGADAQDGLAGLAITDDDMARKAVNYIFETPGGTIYHGADSHFSN 240

Query: 241 YFAKHGKDYKIDVAINNYGDNPNVGIQDKMTSIDLLRMAENLRKVIPVHYDIWSNFMAS 300

YFAKHG+DY IDV +NNYG+NP+GIQDKMTS+DLLRMAENLRKV+IPVHYDIWSNFMAS

Sbjct: 241 YFAKHGRDYDIDVVLNNYGENPIGIQDKMTSVDLLRMAENLRKVIPVHYDIWSNFMAS 300

Query: 301 TDEILQLWKMRRKERLQYDFHPFIWEVGGKYTPQDKDRIEYHHPRGFDDCFEQUESNIQFK 360

TDEIL+LWKMRKERLQYDFHPFIWEVGGKYTPQD++RIEYHHPRGFDDCF ++SNIQFK

Sbjct: 301 TDEILELWKMRRKERLQYDFHPFIWEVGGKYTPQDQNRIEYHHPRGFDDCFLEDSENIQFK 360

Query: 361 ALL 363

ALL

Sbjct: 361 ALL 363

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1505

A DNA sequence (GBSx1592) was identified in *S.agalactiae* <SEQ ID 4629> which encodes the amino acid sequence <SEQ ID 4630>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3988(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10145> which encodes amino acid sequence <SEQ ID 10146> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1662-

>GP:BAA18808 GB:D90917 hypothetical protein [Synechocystis sp.]

Identities = 358/785 (45%), Positives = 494/785 (62%), Gaps = 15/785 (1%)

5 Query: 22 LEKLDAAWRAANYISAQMYLKDNPLLRRELVDLKVHPIGHWGTVPQGNFIYAHNLRA 81
L ++ +WRAANY++ +YL+DNPLLR L +K +GHWG+ PG +F+Y HLNLR
Sbjct: 44 LNQMHGFWRAANYLAVGMIYLRDNPLLRREPLQPEQIKHRLLGHWGSSPGISFLYTHLNRI 103

10 Query: 82 INKYDLDMFYIEGPGHGGQVMVSNYSYLDGSYTELNPNIQTEDGFKQLCKIFSFPFGIAS 141
I K+D DM Y+ GPGHG + YL+GSY+ + EDG K+ K FSFP GI S
Sbjct: 104 IRKFDQDMLYVMVGPFGHAGPGLGPCYLEGSYSRFFAECSEDEDGMRFFKQFSFPGGIGS 163

15 Query: 142 HAAPETPGSIHEGGELGYALSHATGAILDNPDVIAATVIGDGEGETGPLMAGWLSNTFIN 201
H PETPGSIHEGGELGY LSHA GA DNP++I + GDGE ETGPL W SN FIN
Sbjct: 164 HCTPETPGSIHEGGELGYCLSHAYGAADFNPNLIVGLAGDGESETGPLATSWHSNKF 223

20 Query: 202 PVNDGAVLPIFYLNNGKIHNPITIFERKTDEELSQFFEGLGWKPIFADVVELSEDHAAAAHA 261
P+ DGAVLP+ +LNG KI+NP++ R + EEL FEG G+ P F + D + H
Sbjct: 224 PIRDGAVLPVLHLNGYKINNPVLSRISHEELKALFEGYGYTPYFVE---GSDPESMHQ 279

25 Query: 262 LFAEKLDQAIQEIKTIQSEARQKPAEBAIQAKFPVLVARIPKGWTGPKAWEGTPIEGGFR 321
A LD + EI IQ EAR A++ ++P++V R PKGWTGP +G +EG +R
Sbjct: 280 AMAATLDHCVSEIHQIQEARSSTGI--AVRPRWFMVMRTPKGWTGPDYVDGHKVEGFWR 337

30 Query: 322 AHQVPIPVDAHMEHVDLSLWSLQSYRPEELFDENGKIVDEIAAISPKGDRRSMNPITN 381
+HQVP+ + H+ L +W++SY+PEELFDE G + AI+P+GD+R+ P N
Sbjct: 338 SHQVPMGGMHENPAHLQQLAWMSRYKPEELFDEQGTLPKPGFKALAPEGDKRLGSTPYAN 397

35 Query: 382 AGIV-KAMDTADWKKFALDINVPGQIMADMIIEFGKYAADLVDANPDNFRIFGPDET KSN 440
G++ + + D++++ +D++ PG I A + G + D++ N NFR+FGPDE SN
Sbjct: 398 GGLLRRLGLKMPDFRQYIGIDVDQPGTIEAPNTAPLGVFLRDVMANNMTNFRIFGPDENSSN 457

40 Query: 441 RLQEVFTRTSRQWLGRKPKDYDEA--LSPAGRVIDSQLSEHQAEGFLEGYVLTGRHGFFA 498
+L V+ + + W+ + + LSP GRV++ LSEH EG+LE Y+LTGRHGFFA
Sbjct: 458 KLHAVYEVSKKFWIAEYLEEDQDGGELSPDGRVME-MLSEHTLEGWLEAYLLTGRHGFFA 516

45 Query: 499 SYESFLRVDSMVTQHFQWLRKSKTHTTWRKNYPALNLIAASTVFQDHNHGYTHQDPGIL 558
+YESF V+ SMV QH KWL + H WR + +LN++ STV++QDHNH+THQDPG L
Sbjct: 517 TYESFAHVITSMVNQHAKWLDICR-HLNWRADISSLNILMTSTVWRQDHNHGYTHQDPGFL 575

50 Query: 559 THLAETPEYIREYLPADTNSLLAVMDKAFKAEDKINLIVTSKHPRPQFYSAEAEELVA 618
+ K+P+ +R YLP D NSLL+V D +++++ IN+IV K Q+ + A
Sbjct: 576 DVILNKSPDVVRIYLPDPVNSLLSVADHCLQSKNYINIIVCDKQAHLYQYQDMTSAIRNCT 635

55 Query: 619 EGYKVIDWASNVSLNQEPDVVFAAGTEPNLEALAAISILHKAPPELKIRFVNVLDILKL 678
+G + +WASN EPDVP AAAG P EALAA ++L + FP L+IRFV+V+D+LKL
Sbjct: 636 KGVDIWELWASN-DAGTEPDVVMAGADIPTKEALAAATAMLQOFFENLRIRFVSVIDLKL 694

60 Query: 679 RHPSQDARGLSDEEFNKVFTTDKPVIFAFHGYEDMIRDIFFSRHNH-NLHTHGYRENGDI 737
+ S+ GLSD +F+ +FTTDKP+IF FH Y +I + + R NH NLH GY+E G+I
Sbjct: 695 QPESEHPHGLSDRDFSLFTTDKPIIFNFHAYPWLHRLTYRRTNHNHVRGYKEKGNI 754

Query: 738 TTPFDMRVMSLDRFHLAQDA--ALASLGKNAQAFSDENMQMVAYHKDYIREHGDDIPEV 795
TP D+ + +++DRF LA D L L + + M +Y EHG D+PE+
Sbjct: 755 NTPMDLAIQNQIDRFSLAIDVIDRLPQLRVAGAHIKEMLKDMQIDCTNYAYEHGIDMPEI 814

Query: 796 QNWKW 800
NW+W
Sbjct: 815 VNWRW 819

60 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1663-

Example 1506

A DNA sequence (GBSx1593) was identified in *S.agalactiae* <SEQ ID 4631> which encodes the amino acid sequence <SEQ ID 4632>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3509(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF37878 GB:AF234619 OpuAA [Lactococcus lactis]
 Identities = 274/402 (68%), Positives = 338/402 (83%)

Query: 5 LEVKNLTKIFGKKQKAALVMKQGKSKTEILEKTGATVGVDASFEIKEGEIFVIMGLSG 64
 +++++LTKIFGK+ K AL MV++G+ K EIL+KTGATVGVD +FEI EGEIFVIMGLSG
 Sbjct: 5 IKIEHLTKIFGKRIKTALTMVEKGEPKNEILKKTGATVGVDTNFEINEGEIFVIMGLSG 64

Query: 65 SGKSTLVRMLNRLIDPSSGNIYLDGKDIAMNVEDLRNIRRH DINMVFQNFGLFPHTIL 124
 SGKSTL+R+LNRLI+P+SG I++D +D+A +N EDL +RR ++MVFNFGFLFPHTIL
 Sbjct: 65 SGKSTLLRLNRLIEPTSGKIFIDNQDVATLNKEDLLQVRRKTMVMVFQNFGLFPHTIL 124

Query: 125 ENTEFGLEMRGVSKEERTTLAEKALDNAGLLPFKDQYPSQLSGGMQORVGLARALANSFK 184
 ENTE+GLE++ V KEER AEKALDNA LL FKDQYP QLSGGMQORVGLARALAN P+
 Sbjct: 125 ENTEYGLEVQNVKPEERRKRAEKALDNANLLDFKDQYPKQLSGGMQORVGLARALANDPE 184

Query: 185 ILLMDEAFSALDPLIRREMQDELLDLQDTNKTIIIFISHDLNEALRIGDRIALMKDGEIM 244
 ILLMDEAFSALDPLIRREMQDELL+LQ ++TIIF+SHDLNEALRIGDRIA+MKDG+IM
 Sbjct: 185 ILLMDEAFSALDPLIRREMQDELLELQAKFQKTIIFVSHDLNEALRIGDRIAIMKDGKIM 244

Query: 245 QIGTGEEILTNPANDFVREFVEDVDRSKVLT AQNIMIKPLTTVLEIDGPQVALTRMHREE 304
 QIGTGEEILTNPAND+V+ FVEDVDR+KV+TA+NIMI LIT +++DGP VAL +M EE
 Sbjct: 245 QIGTGEEILTNPANDYVKT FVEDVDRKAVITAENIMIPALTTNIDVDGPSVALKKMKTEE 304

Query: 305 VSMLMATNRRRQLLGS LTADAIEARKKDLPLSEVIDKDVTVSKDTVITDIMPLIYDSS 364
 VS LMA +++RQ G +T++ AI ARK + PL +V+ DV TVSK+ ++ DI+P+IYD+
 Sbjct: 305 VSSLMVAVDKKRQFRGVVTSEQAIARKNQPLKDVMTTDVGTVSKEMLVRLDILPIYDAP 364

Query: 365 APIAVTDDNDRLLGVIIIRGRVIEALANVQDET VVESPKETVE 406
 P+AV DDN L GV+IRG V+EALA++ DE VE ++ E
 Sbjct: 365 TPLAVVDDNGFLKGVLRGVSLEALADIPDEDEVVEIEKEEE 406

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4633> which encodes the amino acid sequence <SEQ ID 4634>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3761(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 344/395 (87%), Positives = 374/395 (94%)

Query: 1 MTNILEVKNLTKIFGKKQKAALVMKQGKSKTEILEKTGATVGVDASFEIKEGEIFVIM 60
 M ILEVK+L+KIFGKKQKAALVMK GK+K+EI +KTGATVGVDASFE+K+GEIFVIM
 Sbjct: 1 METILEVKHLSKIFGKKQKAALVMKTKGNKSEIFKKTGATVGVDASFEVKKGEIFVIM 60

Query: 61 GLSGSGKSTLVRMLNRLIDPSSGNIYLDGKDIAMNVEDLRNIRRH DINMVFQNFGLFP 120

-1664-

GLSGSGKSTLVRMLNRLI+PS+G+I L+GKDI+ M+ + LR +RRHDINMVFQ+F LFPH
 Sbjet: 61 GLSGSGKSTLVRMLNRLIEPSAGSILLEGKDISTMSADQLREVRRHDINMVFSFALFPH 120
 Query: 121 RTILENTEFGLEMRGVSKERTTTLAEKALDNAGLLPFKDQYPSQLSGGMQQRVGLARALA 180
 +TILENTEFGLE+RGV KEER LAEKALDN+GLL FKDQYP+QLSGGMQQRVGLARALA
 Sbjet: 121 KTILENTEFGLELRGVPKERQRLAEKALDNSGLLDFKDQYPNQLSGGMQQRVGLARALA 180
 Query: 181 NSPKILLMDEAFSALDPLIRREMDELDDLQDTNKQTIIFISHDLNEALRIGDRIALMKD 240
 NSPKILLMDEAFSALDPLIRREMDELDDLQD+ KQTIIFISHDLNEALRIGDRIALMKD
 Sbjet: 181 NSPKILLMDEAFSALDPLIRREMDELDDLQDSMKQTIIFISHDLNEALRIGDRIALMKD 240
 Query: 241 GEIMQIGTGEEILTNPANDFVREFVEDVDRSKVLTAQNIMIKPLTTTVEIDGPPQVALTRM 300
 G+IMQIGTGEEILTNPANDFVREFVEDVDRSKVLTAQNIMIKPLTT +E+DGPQVAL RM
 Sbjet: 241 GQIMQIGTGEEILTNPANDFVREFVEDVDRSKVLTAQNIMIKPLTTTVELDGPQVALNRM 300
 Query: 301 HREEVSMMLATNRRRQLGSLTADAAIEARKKDLPLSEVIDKDVTVSKDTVITDILPLI 360
 H EEVSMMLATNRRRQL+GSLTADAAIEARKK LPLSEVID+DV TVSKDT+ITDI+PLI
 Sbjet: 301 HNEEVSMMLATNRRRQLVGSLTADAAIEARKKGLPLSEVIDRVRTVSKDTIITDILPLI 360
 Query: 361 YDSSAPIAVTDDNDRLLGVIIIRGRVIEALANVODE 395
 YDSSAPIAVTDDN+RLLGVIIIRGRVIEALAN+ DE
 Sbjet: 361 YDSSAPIAVTDDNDRLLGVIIIRGRVIEALANISDE 395

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1507

A DNA sequence (GBSx1594) was identified in *S.agalactiae* <SEQ ID 4635> which encodes the amino acid sequence <SEQ ID 4636>. This protein is predicted to be OpuABC (opuAB). Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -10.67 Transmembrane 48 - 64 (43 - 72)
 INTEGRAL Likelihood = -9.24 Transmembrane 101 - 117 (93 - 122)
 INTEGRAL Likelihood = -7.54 Transmembrane 296 - 312 (290 - 316)
 INTEGRAL Likelihood = -6.21 Transmembrane 252 - 268 (250 - 273)
 INTEGRAL Likelihood = -5.57 Transmembrane 141 - 157 (138 - 170)
 INTEGRAL Likelihood = -0.53 Transmembrane 220 - 236 (220 - 237)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5267(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF37879 GB:AF234619 OpuABC [Lactococcus lactis]
 Identities = 345/578 (59%), Positives = 429/578 (73%), Gaps = 8/578 (1%)
 Query: 1 MENLLQHKLPVAPFVESTTNWITKTFSGLFDFIQTIGNALMDWMTKTLLFINPLLFIVLI 60
 M +L ++P+A +V S T+WIT TFS FD IQ G LM+ +T L + L I ++
 Sbjet: 1 MIDLAIGQVPIANWVSSATDWITSTFSSGFDVIQKSGTVLMNGITGALTAVPFWLMIAVV 60
 Query: 61 TIAVFFLAKKKWQLPTFTFIGLLFIYNQGLWEQLINTFNLVLVASLISIIIGVPLGIWMA 120
 TI ++ KK P FTFIGL I NQGLW L++T LVL++SL+SIIGVPLGIWMA
 Sbjet: 61 TILAILVSGKKIAFPFLFTFIGLSLIANQGLWSDLMSTITLVLSSLSIIIGVPLGIWMA 120
 Query: 121 KSDKVKQVNPILDFMQTMPAFVYLIPAVAFFGIGMVGPFASVVFALPPTVRFTNLAIR 180
 KSD V ++V PILDFMQTMP FVYLIPAVAFFGIG+VPGVFASV+FALPPTVR TNL IR
 Sbjet: 121 KSDLVAKIVQPIILDFMQTMPGFVYLIPAVAFFGIGVVPVFASVIFALPPTVRMTNLGIR 180
 Query: 181 EIPLELIEASDSFGSTVKQKLFKVELFLAKNTIMAGINQTMMLALSMVVTGSMIGAPGLG 240
 ++ EL+EA+DSFGST +QKLFK+E FLAK TIMAG+NQT+MLALSMVV SMIGAPGLG

-1665-

5 Sbjct: 181 QVSTELVEADSFSGSTARQKLFKLEFPLAKGTIMAGVNQTIMLALSMVVIASMGAPGLG 240
 Query: 241 REVLALQHADIGTGFVSGLSLVILAIVLDRVSQFFNSKPGKQAKTSKVKKW--VGLG 297
 R VL+A+Q ADIG GFVSG+SLVILAI++DR +Q N P EKQ + VKKW + L
 10 Sbjct: 241 RGVLAAVQSADIGKGFVSGISLVILAIIDRFTQKLNVSPLKQGNPT-VKKWKRGIALV 299
 Query: 298 ALALFILAAALGRIVVNMSTSGNEAKGQKVKIAYVQWDSEVASTNVIAEVLKSKGYDVELTP 357
 +L I+ A M+ G A +KV+ Y+ WDSEVAS NV+ + +K G+DV+ T
 15 Sbjct: 300 SLLALIIGAFS----GMSFGKTASDKKVDLVYMNWDSEVASINVLTQAMKEHGFVDVKTTA 355
 Query: 358 LDNAVWQTVANGNADFTTSAWLPKTHGQYFNKYKNSLDDLGPVENVKIGLVVPKYMNV 417
 LDNAV WQTVANG AD SAWLP TH + KY S+D LGP+++ K+G VVP YMN
 20 Sbjct: 356 LDNAVAWQTVANGQADGMVSAWLPNTHKTQWQKYKGSVDLLGPNLKGAKEVGFVVPYMN 415
 Query: 418 NSIEELSNAQADKQITGIEPGAGIMKSAKQSLKDYENLSSWKLSSASTGAMTTTLGKAIKN 477
 NSIE+L+NQA+K ITGIEPGAG+M ++++L Y NL WKL+ +S+GAMT LG+AIK
 25 Sbjct: 416 NSIEDLTNAQANKTITGIEPGAGVMAASEKTLNSYDNLKDWKLPSSSGAMTVALGEAIKQ 475
 Query: 478 KDQVVITGWSPHWMFAKYDLKYLKDPKKSFGGEEHINTIARKNLKDKMPKVYKIIDKFKW 537
 +VITGWSPHWMF KYDLKYL DPK + G E+INTI RK LKK+ P+ YK++DKF W
 30 Sbjct: 476 HKDIVITGWSPHWMFNKYDLKYLADPKGTMTSENINTIVRGLKKENPEAYKVLDKFNW 535
 Query: 538 TKEDMESIMLMDKGMPEAKAAQKWKIKNHKKEVSEWTK 575
 T +DME++MLD+ G P +AA+ WIK+H+KEV +W K
 35 Sbjct: 536 TTKDMEAVMLDIQNGKTPEEAAKNWIKDHQKEVDKWF 573

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4637> which encodes the amino acid sequence <SEQ ID 4638>. Analysis of this protein sequence reveals the following:

30 Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.86 Transmembrane 101 - 117 (93 - 121)
 INTEGRAL Likelihood = -7.54 Transmembrane 252 - 268 (250 - 273)
 INTEGRAL Likelihood = -6.85 Transmembrane 48 - 64 (43 - 70)
 35 INTEGRAL Likelihood = -5.57 Transmembrane 141 - 157 (138 - 170)
 INTEGRAL Likelihood = -5.26 Transmembrane 295 - 311 (289 - 315)
 INTEGRAL Likelihood = -0.53 Transmembrane 220 - 236 (220 - 237)
 ----- Final Results -----
 40 bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 >GP:AAF37879 GB:AF234619 OpuABC [Lactococcus lactis]
 Identities = 340/571 (59%), Positives = 418/571 (72%), Gaps = 8/571 (1%)
 Query: 8 KLPVAQLVEQLTEWLTKTFSGLEFDIMQVVSFLMDWMTKTLFLFIHPLLFIVLVTAGMFPL 67
 ++P+A V T+W+T TFS FD++Q G+ LM+ +T L + L I +VT +
 50 Sbjct: 8 QVPIANWVSSATDWITSTFSSGFDVIQSGTVLMNGITGALTAVPFWLMIAVVTILAILV 67
 Query: 68 AKKKWPLPTFTLLGLLFIYNQGLWQLMNTFTLVLVASLISVLIGIPLGIWMAKNATVRQ 127
 + KK P FT +GL I NQGLW LM+T TLVL++SL+S++IG+PLGIWMAK+ V +
 55 Sbjct: 68 SGKKIAPPLFTFTIGLSLIANQGLWSDLMSTITLVLLSSLSIIIGVPLGIWMAKSDLVAK 127
 Query: 128 IVNPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASVIFALPPTVRFTNLAIIRDIPTELI 187
 IV PILDFMQTMP FVYLIPAVAFFGIG+VPGVFASVIFALPPTVR TNL IR + TEL+
 60 Sbjct: 128 IVQPIILDFMQTMPGFVYLIPAVAFFGIGVVPGVFASVIFALPPTVRMTNLGIRQVSTELV 187
 Query: 188 EASDAFGSTGKQKLFKVELPLAKNTIMAGVNQTMMLALSMVVTGSMIGAPGLGREVLAL 247
 EA+D+FGST +QKLFK+E PLAK TIMAGVNQ+MLALSMV SMIGAPGLGR VL+A+
 65 Sbjct: 188 EAADSFSGSTARQKLFKLEFPLAKGTIMAGVNQTIMLALSMVVIASMGAPGLGRGVLA 247
 Query: 248 QHADIGSGFVSGLALVILAIVLDRMTQLFNSKPQEKAKAGKTNKW---IGLAALAVFLIA 304
 Q ADIG GFVSG++LVILAI++DR TQ N P EK KW I L +L +I
 70 Sbjct: 248 QSADIGKGFVSGISLVILAIIDRFTQKLNVSPLKQGNPTVKKWKRGIALVSLALIIG 307

-1666-

5 Query: 305 ALGRGIMAMTSGMADKGETVNIAYVQWDSEVASTHVIAEVLKNEGYHVTITPLDNAVMWQ 364
 A M+ G + V++ Y+ WDSEVAS +V+ + +K G+ V T LDNAV WQ
 Sbjct: 308 AFS----GMSFGKTASDKKVDLVYMNWDSEVASINVLTOAMKEHGFDVKTALDNAVAWQ 363

10 Query: 365 TVANGNADFSTSAWLPTVTHGQQYQKYKSLDDLGPNLKGTGLGLAVPKYMTDVNSIEDLS 424
 TVANG AD SAWLP TH Q+QKY +D LGPNLKG K+G VP YM +VNSIEDL+
 Sbjct: 364 TVANGQADGMVSAWLPNTHKTQWQKYGKSVDDLGPNLKGAKVGFVPSYM-NVNSIEDLT 422

15 Query: 425 KQADQKITGIEPGAGIMAAAQKTLKEYHNLSSWELVAASTGAMTSLDQAIKKKDPVIVT 484
 QA++ ITGIEPGAG+MAA++KTL Y NL W+LV +S+GAMT +L +AIK+ IV+T
 Sbjct: 423 NQANKTITGIEPGAGVMAASEKTLNSYDNLKDWKLVPSSSGAMTVALGEAIKQHKDIVIT 482

20 Query: 485 AWSPHWMFAKYDLKYLKDPKEIFGSTENINTIARKGLKKELPNVYKIIDKFHWTKDMEA 544
 WSPHWMF KYDLKYL DPK G++ENINTI RKGLKKE P YK++DKF+WT KDMEA
 Sbjct: 483 GWSPHWMFNKYDLKYLADPKGTMGTSNINTIVRKGLKKENPEAYKVLDFKNWTTKDMEA 542

Query: 545 VMLDINKGMSPEAAAKKWEANKSKVSSWTK 575
 VMLDI G +PE AAK W++ ++ +V W K
 Sbjct: 543 VMLDIQNGKTPEEAAKNWIKDHQKEVDKWF 573

An alignment of the GAS and GBS proteins is shown below.

Identities = 439/576 (76%), Positives = 513/576 (88%), Gaps = 2/576 (0%)

25 Query: 1 MENLLQHKLPVAPFVESTTNWITKTFSGLFDIFIQTIGNALMDWMTKTLFINPLLFIIVLI 60
 +E +LQ KLPVA VE T W+TKTFSGLFD +Q +G+ LMDWMTKTLFI+PLLFIIVL+
 Sbjct: 1 LETILQTKLPVAQLVEQLTEWLTKTFSGLFDIMQVGSFLMDWMTKTLFIHPLLFIIVL 60

30 Query: 61 TIAVFFLAKKKWQLPTFTFIGLLFIYNQGLWEQLINTFNLVLVASLISIIIGVPLGIWMA 120
 T +FFLAKKKW LPTFT +GLLFIYNQGLW+QL+NTF LVLVASLIS++IG+PLGIWMA
 Sbjct: 61 TAGMFFLAKKKWPLPTFTLLGLLFIYNQGLWQLMNTFTLVLVASLISVLIGIPLGIWMA 120

35 Query: 121 KSDKVQVVPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASVVFALPPTVRFTNLAIR 180
 K+ V+Q+VNPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASV+VFPPTVRFTNLAIR
 Sbjct: 121 KNATVRQIVNPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASVIFALPPTVRFTNLAIR 180

40 Query: 181 EIPLELIEASDSFGSTVKQLFKVELPLAKNTIMAGINQTMMLALSMVVTGSMIGAPGLG 240
 +IP ELIEASD+FGST KQKLFKVELPLAKNTIMAG+NQTMMLALSMVVTGSMIGAPGLG
 Sbjct: 181 DIPTLELIEASDAFGSTGKQLFKVELPLAKNTIMAGVNQTMMLALSMVVTGSMIGAPGLG 240

45 Query: 241 REVLALQHADIGTGFVSGLSVLILAIVLDRVSQFFNSKPGEKQAKTSKVKKWVGLGALA 300
 REVLALQHADIG+GFVSGL+LVILAIVLDR++Q FNSKP EK AK K KW+GL ALA
 Sbjct: 241 REVLALQHADIGSGFVSGLAIVILAIVLDRMTQLFNSKPQEK-AKAGKTNKWIGLAALA 299

50 Query: 301 LFILAALGRIVVNMSTGNEAKGQKVIAYVQWDSEVASTNVIAEVLKSGYDVELTPLDN 360
 +F++AALGR ++ MTSG KG+ V IAYVQWDSEVAST+VIAEVLK++GY V LTPLDN
 Sbjct: 300 VFLIAALGRGIMAMTSGMADKGETVNIAYVQWDSEVASTHVIAEVLKNEGYHVTITPLDN 359

55 Query: 361 AVMWQTVANGNADFSTSAWLPTVTHGQQYFNKYKNSLDDLGPVENVKIGLVVPKYM-NVNS 419
 AVMWQTVANGNADF+TSAWLP THGQ + KYK+ LDDLGP+++ K+GL VPKYM +VNS
 Sbjct: 360 AVMWQTVANGNADFSTSAWLPTVTHGQQYQKYKSLDDLGPNLKGTGLGLAVPKYMTDVNS 419

60 Query: 420 IEELSNOADKQITGIEPGAGIMSAKQSLKDYPNLSSWKLLSASTGAMTTTLGKAIKND 479
 IE+LS QAD++ITGIEPGAGIM +A+++LK+Y NLSSW+L++ASTGAMTT+L +AIK KD
 Sbjct: 420 IEDLSQADKQITGIEPGAGIMAAAQKTLKEYHNLSSWELVAASTGAMTSLDQAIKKD 479

65 Query: 480 QVVITGWSPHWMFAKYDLKYLKDPKKSFGGEEHINTIARKNLKDKMPKVYKIIDKFKWT 539
 +V+T WSPHWMFAKYDLKYLKDPK+ FG E+INTIARK LKK++P VYKIIDKF WT+
 Sbjct: 480 PIVVITAWSPHWMFAKYDLKYLKDPKEIFGSTENINTIARKGLKKELPNVYKIIDKFHWQ 539

Query: 540 EDMESIMLMDMDKGMEPAKAAQKNHKKVESEWTK 575
 +DME++MLD++KGM P AA+KW++ +K +VS WTK
 Sbjct: 540 KDMEAVMLDINKGMSPEAAAKKWEANKSKVSSWTK 575

65 A related GBS gene <SEQ ID 8827> and protein <SEQ ID 8828> were also identified. Analysis of this protein sequence reveals the following:

-1668-

```

|||: :||| ||: ||| :||| ||| ||| ||| ||| : | |: ||| ||| :| :|||
DWKLVPSSESGAMTVLGEATKQHKDITVTGWSFHWMFNFKYDLKYLADPKGTMTSENINTIVRKGLKKENPEAYKVLDFK
470      480      490      500      510      520      530

5  1935      1965      1995      2025      2055      2085      2115      2145
   KWTKEDESMESIMLDMDKGMPEAKAAQKWKIKNHKKVSEWTK*YRKKHVSFRACFLM*LKSF*LFNISFILF*YIKSERMKE
   || :|||: |||: | | :||: |||: |||: ||| :| |
   NWTTKDMEAVMLDIQNGKTPFEAAKNWIKDHQKEVDKWKFK
10      550      560      570

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1508

A DNA sequence (GBSx1596) was identified in *S. agalactiae* <SEQ ID 4639> which encodes the amino acid sequence <SEQ ID 4640>. This protein is predicted to be a transposase. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -1.65    Transmembrane  223 - 239 ( 223 - 240)

----- Final Results -----
      bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10057> which encodes amino acid sequence <SEQ ID 10058> was also identified. A related GBS nucleic acid sequence <SEQ ID 10031> which encodes amino acid sequence <SEQ ID 10032> was also identified. A related GBS nucleic acid sequence <SEQ ID 10801> which encodes amino acid sequence <SEQ ID 10802> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA50689 GB:X71844 putative transposase [Clostridium perfringens]
Identities = 94/364 (25%), Positives = 160/364 (43%), Gaps = 35/364 (9%)

Query: 8   KHKHLLTLDRNDIQSGLDRGETFKAIGLNLKHPPTTIAKEVKRN--KQLRESTKDCLDCP 65
          K+KHL + +R ++ L G +          L + T+ E++R KQ+++ + +
Sbjct: 12  KNKHLNMKERMIVEIRLKDGFSAKYNTKELNRPINTVLNEIRRGTTKQIKQGKEFHVYFA 71

Query: 66  LLRKAPYVCNGCPKRRINCQYKKTFFYLAKQAQRNYEKLIVESREGIPLNKETFWKIDRVL 125
          +A Y N + + N YK          ++ K +V+          K W +D +
Sbjct: 72  DTGEAVYKKN---RLKSNRKYKLL-----ECSDFIKYVVDKV-----KNDHWSLDACV 116

Query: 126 SNGVKKGQRIYHILKTNDLEVSSSTVYRHIKKGYSITPIDLERAVKFKKRRKSTLPPIP 185
          G+ ++ + +S+ T+Y ++ G L I IDLP K + +KST
Sbjct: 117 -----GEALHSSRFSPSQIISTKTLYNYVDLGLLPKIDLP--AKLHRNKKSTRVRNN 168

Query: 186 KAIKEGRRYEDFIEHM-NQSELNSWLEMDTVIGRIGGK--VLLTFNVAFCNFIFAKLMDS 242
          K K G D + N+ E W E+D V+G K VLLT + M S
Sbjct: 169 KK-KLGTSISDRPNISIENREEFGHW-EIDCVLGEKSNKDKVLLTLVERKTRYAISEMSS 226

Query: 243 KTAIETAKHIQVIKRTLYDNKRDFELFPVILTNGGEFARVDDIEIDVCGQSOLFCDP 302
          + I K + IK L F E+F I DNG EFA + + E+ +++++F P
Sbjct: 227 HSTISVTKALDKIKEFLGSK---FSEVFKSITADNGSEFADLSEFELKT--KTKVYFTHP 281

Query: 303 NRSDQKARIEKNHTLVRDILPKGTSFDNLTQEDINLALSHINSVKRQALNGKTAYELFSF 362
          S +K E+++ L+R +PKG + + E I+ + +N++ R+ L+ KT ELF
Sbjct: 282 YSSFEKGTNERHNGLIIRRFIPKGRISDYSLETISFIENWMNTLPRKLLDYKTPEELFEI 341

Query: 363 TYGK 366

```

-1669-

K
Sbjct: 342 HLDK 345

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1509

A DNA sequence (GBSx1597) was identified in *S.agalactiae* <SEQ ID 4641> which encodes the amino acid sequence <SEQ ID 4642>. Analysis of this protein sequence reveals the following:

10 Possible site: 33
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -11.30	Transmembrane	56 - 72 (48 - 79)
INTEGRAL	Likelihood = -6.85	Transmembrane	11 - 27 (6 - 30)
INTEGRAL	Likelihood = -6.69	Transmembrane	129 - 145 (126 - 158)
15 INTEGRAL	Likelihood = -6.53	Transmembrane	94 - 110 (90 - 117)
INTEGRAL	Likelihood = -1.54	Transmembrane	216 - 232 (215 - 232)
INTEGRAL	Likelihood = -1.22	Transmembrane	147 - 163 (147 - 165)

----- Final Results -----

20 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 9431> which encodes amino acid sequence <SEQ ID 9432>
25 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:EAB07666 GB:AP001520 unknown conserved protein [Bacillus halodurans]
Identities = 112/224 (50%), Positives = 150/224 (66%), Gaps = 2/224 (0%)

30 Query: 8 IKDILWFIIPSLFGVLLMTFPKYNGMTTAVSVISKTNQWINAVFPIHYIILLIIFIS 67
+KD LWF+IPS+ GV L M P + + T+ V+ ++K + ++ P I+L I +
Sbjct: 19 LKDYLWFLIPSIIGVGLFMVPIQKDNATIPVAFLAKLQGLDDHLPAILTIMLAIVV- 77

35 Query: 68 CVLALCYRLFRPSFIEKNDLLKEISDITIFWLIIRLIGLALGLMTVLHIGPEMVWGKETG 127
VL+ LF+P+ KN LLK + I WL++R++G MT+L +GPE VW + TG
Sbjct: 78 -VLSCVATLFKPNLFMKNGLKSLFVIHPMWLVVRVLGFIFAFMTLLQLGPEAVWSEGTG 136

Query: 128 GLILFDLIGGLFTIFLAAGFILPFLTEFGLLEFVGVLTPIMRPPFQLPGRSAVNCVASF 187
L+L+DL+ LFTIFL AG LPFL FGLLE GV L MRP F LPGRS+++C+AS+
40 Sbjct: 137 ALLLYDLLPLLFITIFLAFGLFLPFLNLFGLLEFGLVLLNKFMRPVFTLPGRSSIDCLASW 196

Query: 188 VGDGTIGIALTDKQYVEGYYSREAAITSTTFSVSTFCLXXL 231
+GDGTIG+ LT+KQY EG+YT REAA ISTTFS VSITF + L
45 Sbjct: 197 MGDGTIGVLLTNKQYEEGFYQREAAVISTTFSVVSITFSIVVL 240

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1510

- 50 A DNA sequence (GBSx1599) was identified in *S.agalactiae* <SEQ ID 4643> which encodes the amino acid sequence <SEQ ID 4644>. This protein is predicted to be Na/H antiporter homolog (kefB). Analysis of this protein sequence reveals the following:

-1670-

Possible site: 17

>>> Seems to have an uncleavable N-term signal seq

	INTEGRAL	Likelihood = -10.14	Transmembrane	176 - 192 (171 - 203)
	INTEGRAL	Likelihood = -9.34	Transmembrane	353 - 369 (348 - 373)
5	INTEGRAL	Likelihood = -9.24	Transmembrane	3 - 19 (1 - 26)
	INTEGRAL	Likelihood = -7.17	Transmembrane	145 - 161 (142 - 168)
	INTEGRAL	Likelihood = -7.01	Transmembrane	86 - 102 (81 - 108)
	INTEGRAL	Likelihood = -6.53	Transmembrane	52 - 68 (51 - 72)
	INTEGRAL	Likelihood = -5.79	Transmembrane	24 - 40 (23 - 49)
10	INTEGRAL	Likelihood = -5.52	Transmembrane	214 - 230 (209 - 233)
	INTEGRAL	Likelihood = -4.04	Transmembrane	260 - 276 (258 - 278)
	INTEGRAL	Likelihood = -3.66	Transmembrane	287 - 303 (287 - 308)
	INTEGRAL	Likelihood = -2.71	Transmembrane	113 - 129 (112 - 129)
15	INTEGRAL	Likelihood = -2.66	Transmembrane	332 - 348 (330 - 349)

----- Final Results -----

	bacterial membrane	---	Certainty=0.5055(Affirmative)	< succ>
	bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
20	bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA51756 GB:X73329 Na/H antiporter homolog [Lactococcus lactis]

Identities = 208/376 (55%), Positives = 285/376 (75%), Gaps = 3/376 (0%)

25	Query: 1	MHIIQITIIILLASVLATLISKRIGIPAVVGQQLLVGIIIGPAMLGLVHQNQVLHVLSEIG 60
		M+ I+Q+TI+L+AS++ATL S+R+ IPAV+GQ+LVGI+I P++LGLVH VL V+SEIG
	Sbjct: 1	MNDILQLTIVLIASLIATLASRRLLKIPAVIGQMLVGLIAPSVLGLVHSGHVLEVMSEIG 60
30	Query: 61	VILLMFLAGLEANFDLLKKYLKPSLLVAITGVIVPMALFYFLTRLFGFQINTAIFYGLVF 120
		VILLMFLAGLE++ +LKK K S+LVAI GVIVP+ +F + FG+ ++T+ FYG+VF
	Sbjct: 61	VILLMFLAGLESDLTVLKKNFKASMLVAIGGVIVPLIVFGLVAFSFGYGMSTSFYFYGIVF 120
35	Query: 121	AATSISITVEVLQEQYNRVKTDGTGAILGAAVADDVLAVLLLSVFIA--TNGSSSNIGLQI 178
		AATS+SITVEVLQEQY ++ T G+IILGAAV DD+LAVL+LS+F + GS +++ Q
	Sbjct: 121	AATSVSITVEVLQEQYGLSTRAGSIILGAAVDDILAVLILSIFTSFKNGSGTHLFFQF 180
40	Query: 179	IIQLLEFFVPLFICMKYLPALFKLIEKVHFFFEKYTILAILICFSLILADKVGMSIIGS 238
		+++LLFF FLF+ K L+P +K ++K+ K TI+A++IC LS+LAD VGMS++IGS
	Sbjct: 181	LLELLFFAFLEFVVHK-LIPRFWKVFQKLPANKNITIVALIICLGLSLLADSVGMSAVIGS 239
45	Query: 239	FFAGLAIGQTSFVDKVEHKISLLSYTFFPIPIFFASIALPLKFDGMMSHLHTILIFTALAV 298
		FFAGLAI QT K+E S + Y FIP+FF IA+ ++FD ++ H IL+FT LA+
	Sbjct: 240	FFAGLAISQTEVSHKIEEYTSAGYVIFIPVFFVLIAISVQFDSLIIHPWIILFTLLAI 299
50	Query: 299	LSKLIPGYFVGRGFNFSLKESITIGGGMVSRGEMALIIVQVGLAAKIISSTTYSELVIVV 358
		L+K IP YFVG+ S ES+ IG GM+SRGEMALI+ Q+GL + II+ YSELVIV+
	Sbjct: 300	LTKFIPAYFVGKSNKLSTGESMLIGTGMISRGEMALIVAQIGLTSAITTDEVYSELVIVI 359
	Query: 359	ILSTIIAPFILKYSFK 374
		IL+T++APF++K K
	Sbjct: 360	ILATVLAPFLIKLVLK 375

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1511

A DNA sequence (GBSx1600) was identified in *S.agalactiae* <SEQ ID 4645> which encodes the amino acid sequence <SEQ ID 4646>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

-1671-

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14269 GB:Z99116 ypuA [Bacillus subtilis]
 Identities = 86/319 (26%), Positives = 147/319 (45%), Gaps = 34/319 (10%)

10

Query: 3 IKKLLFAGLAFILFTLASPAYAASDVQKVIDETVQPDYVLGYSLNQEQRAQTQLLLNYD 62
 +KK+ LA + L P + +D + + V LG L++ + + L +N
 Sbjct: 1 MKKIWIGMLAAAVLLLMVPKVSLEDA--AVGDVIV---TLGADLSESDKQKVLDEMNPV 54

15

Query: 63 ESRDTKVKTLNTSSYAKIMNIADDASIQLY---SSVKIKKLGSNDTLAVNIVTPENITK 118
 ++ T V N + + +A I SS+ I K GS +N+ T NI+
 Sbjct: 55 DNAIT-TVTINKKEHEYLKGYISNAQIGSRAISSSSITIAKKGSG---LNVET-HNISC 108

20

Query: 119 VTEDMYRNAAVTLGIEHATISVAAPIKVTGESALAGIYYSLE-KNGASVSSENKQLAQEE 177
 +T++MY NA +T G++ A + V AP +V+G +AL G+ + E + ++S + KQ+A +E
 Sbjct: 109 ITDEMYLNALMTAGVKDAKVYVTAPFEVSGTAALTGLIKAYEVSSDEAISEDVKQVANQE 168

25

Query: 178 LSTLSGINAENKGKEGYDADKLNVALTDIKSAVAKGGSDDLKDDIRKIVEETLKNYHLDN 237
 L T S + + G E A + I K AK G +K DI K V++ + L+
 Sbjct: 169 LVTTSEL-GDKIGNENAAA-----LIAKIKEEFAKNGVPDNKADIEKQVDDAASD--LNV 220

30

Query: 238 AVTENQINLIVNFAVNLSQSNNVKNSDFTNTLNNLKDNIVSKAGSKFKNINVNFNANKAV 297
 +T++Q N +V S N +KN+D + D + KA K + +
 Sbjct: 221 TLTDQKQNLV-----SLFNKMKNADI--DWGQVSDQL-DKAKDKITKFIESDEGKNFI 271

Query: 298 ESGKGFLANIWQQIVNFFQ 316
 + F +IW IV+ F+
 Sbjct: 272 QKVIDFFVSIWNAIVSIFK 290

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1512

A repeated DNA sequence (GBSx1602) was identified in *S.agalactiae* <SEQ ID 4647> which encodes the amino acid sequence <SEQ ID 4648>. Analysis of this protein sequence reveals the following:

40

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45

bacterial cytoplasm --- Certainty=0.0603(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50

>GP:CAB15719 GB:Z99122 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 76/138 (55%), Positives = 91/138 (65%), Gaps = 12/138 (8%)

55

Query: 1 MKLKAVHHIAIIVSDYEKSKDFYVKNLGFELIIRENHRPERHDKLDLRC-GDIELEIFGN 59
 M LK++HHIAII SDYEKSK FYV+KLG++I+E +R ER YKLDL G +E+F
 Sbjct: 1 MLLKSIHHIAIICSDEYKSKAFYVHKLGFQVIQETTYREERGSYKLDLSLNGSYVIELF-- 58

Query: 60 RLDDPEYETPPQIRGRPNWPREACGLRHLAFYVPDVEAYKVELENLGI FVEPIRYDDYTG 119
 + PP+R RP EA GLRHLAF V ++ EL GI EPIR D TG
 Sbjct: 59 -----SFPDPPERQTRP-----EAAGLRHLAFTVGSGLDKAVQELHEKGIETEPITRDPLTG 109

-1672-

Query: 120 KMTFFFDPDGLPLELHE 137
 K+ TFFFDPD LPLEL+E
 Sbjct: 110 KRFTFFFDPDQLPLELYE 127

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4649> which encodes the amino acid sequence <SEQ ID 4650>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1205(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 99/137 (72%), Positives = 116/137 (84%)

Query: 1 MKLKAVHHIAIIVSDYEKSKDFYVKNLGFETIIRENHRPERHDYKLDLRCCDIELEIFGNR 60
 MKL A+HH+AIIVSDY SKDFYVKNLGFETIIREN+RP++HDYKLDL CG IELEIFG
 20 Sbjct: 2 MKLNIAHHVAIIVSDYHLSKDFYVKNLGFETIIRENRYRDPKHDKLDLSCGRIELEIFGKV 61
 Query: 61 LDDPEYETPPQIRGNPNWPREACGLRHLAFYVPDVEAYKVELENLGI FVEPIRYDDYT GK 120
 DP Y+ PP+R+ P + EACGLRHLAF V ++E+Y +L++LGI VEPIR+DDYT G+
 Sbjct: 62 TSDPNYQAPPKRVSSEPEFKSEACGLRHLAFRVNIESYVDDLKSLGIPVEPIRHDDYTGE 121
 25 Query: 121 KMTFFFDPDGLPLELHE 137
 KMTFFFDPDGLPLELHE
 Sbjct: 122 KMTFFFDPDGLPLELHE 138

- 30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1513

- A DNA sequence (GBSx1603) was identified in *S.agalactiae* <SEQ ID 4651> which encodes the amino acid sequence <SEQ ID 4652>. This protein is predicted to be alpha-amylase. Analysis of this protein
 35 sequence reveals the following:

Possible site: 40
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -11.62 Transmembrane 14 - 30 (7 - 36)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG41778 GB:AF213261 sortase [Streptococcus gordonii]
 Identities = 136/247 (55%), Positives = 174/247 (70%), Gaps = 2/247 (0%)

Query: 2 RNKKKSHGFFNFVRWLLVLLIIVGLALVFNKPIRNAFIAHQSNHYQISRVSKKTIEKNK 61
 R KK N + +L V+L++V LAL+FN IRN + +N YQ+S+VSKK IEKNK
 50 Sbjct: 6 RRAKKRSRRNIILNLSVILLLVALALIFNSSIRNMIMVWHTNKYQVSKVSKKEIEKNK 65
 Query: 62 KSKTSYDFSSVKSISTESILSAQTKSHNLPVIGGIAIPDVEINLPFKGLGNTLSYGAG 121
 SK S++F V+ +STE++L+AQ K+ LPVIGGIAIP++ +NLPIF GL N L YGAG
 55 Sbjct: 66 ASKGSFNFEKVEPLSTEAVLNAQWKAQQLPVIGGIAIPELSLNLPIFNGLNAGLYYGAG 125
 Query: 122 TMKENQIMGGPNNYALASHHVFLTGSSKMLFSPLEHAKKGMKVYLTDKSKVYTYTTITEI 181
 TMKE Q M G NYALASHHVFG+TG+++MLFSPL+ AK GMK+YLTDK KVTY+IT +

-1673-

Sbjct: 126 TMKETQEM-GKGNALASHHVFGITGANEMLFSPDRAGMKIYLTDKKQVYTSITSV 184

Query: 182 SKVTPHEHVEIDD-TPGKSQTLTVCTDPEATERIIVHAELEKTGEFSTADESILKAFSK 240
V PE V+V+DD G +++TLVTC D AT R IV LE + + IL F+K

Sbjct: 185 ENVEPERVDVDDAADGTAEVTLVTCEDAAATSRTIVKGVLESETPYKETPKKILNYFNK 244

Query: 241 KYNQINL 247

YNQ+ L

Sbjct: 245 SYNQMQL 251

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4653> which encodes the amino acid sequence <SEQ ID 4654>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.12 Transmembrane 18 - 34 (13 - 38)

INTEGRAL Likelihood = -0.32 Transmembrane 94 - 110 (94 - 110)

----- Final Results -----

bacterial membrane --- Certainty=0.4248(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAA73122 GB:M77279 alpha-amylase [unidentified cloning vector]

Identities = 60/122 (49%), Positives = 85/122 (69%)

Query: 7 RRKIKMSWARKLLIAVLLILGLALLFNKPIRNTLIARNSNKYQVTKVSKKQIKKNKEAKS 67
+ K + +W L+ +L I+GLAL+FN IR+ ++ +NS Y V+K+ +KKN ++

Sbjct: 4 KEKRGKNWNLINSLVLLFIIGLALIFNNQIRSWVQQNSRSYAVSKLKPADVKKNMARET 64

Query: 68 TFDQAVEPVSTESVLQAQMAAQQLPVIGGIAIPELGINLPIFKGLGNTIELIYGAGTMKEE 127
TFDF +VE +STE+V++AQ + LPVIG IAIP + INLPIFKGL N L+ GAGTMKE+

Sbjct: 65 TFDSDVESLSTEAVMKAQFENKNLPVIGAIAPISVEINLPIFKGLSNVALLTGAGTMKED 124

An alignment of the GAS and GBS proteins is shown below.

Identities = 147/245 (60%), Positives = 192/245 (78%)

Query: 2 RNKKKSHGFFNFVRWLLVLLIIVGLALVFNKPIRNFIAHQSNHYQISRVSKKTIEKNK 61
+ K++ ++ R LL+ +L+I+GLAL+FNKPIRN IA SN YQ+++VSKK I+KNK

Sbjct: 4 KQRRKIKMSWARKLLIAVLLILGLALLFNKPIRNTLIARNSNKYQVTKVSKKQIKKNK 63

Query: 62 KSKTSYDFSSVKSISTESILSAQTKSHNLPVIGGIAIPDVEINLPIFKGLGNTIELSYGAG 121
++K+++DF +V+ +STES+L AQ + LPVIGGIAIP++ INLPIFKGLGNTIEL YGAG

Sbjct: 64 EAKSTFDQAVEPVSTESVLQAQMAAQQLPVIGGIAIPELGINLPIFKGLGNTIELIYGAG 123

Query: 122 TMKENQIMGGPNNYALASHHVFGITGSSKMLFSPLEHAKKGMKVYLTDKSKVYTYTITEI 181
TMKE Q+MGG NNY+LASHH+FG+TGSS+MLFSELE A+ GM +YLTDK K+Y Y I ++

Sbjct: 124 TMKEEQVMGGENNYSLASHHIFGITGSSQMLFSPLEAQNMGMSIYLTDKKEKIYEYI IKDV 183

Query: 182 SKVTPHEHVEIDTPGKSQTLTVCTDPEATERIIVHAELEKTGEFSTADESILKAFSKK 241
V PE V+VIDDT G ++TLVCTD EATERIIV EL+ +F A +LKAF+

Sbjct: 184 FTVAPERVDVIDDTAGLKEVTLVCTDIEATERIIVKGELKTEYDFDKAPADVLKAFNHS 243

Query: 242 YNQIN 246

YNQ++

Sbjct: 244 YNQVS 248

SEQ ID 4652 (GBS266) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 11; MW 26kDa).

GBS266-His was purified as shown in Figure 205, lane 10.

-1674-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1514

A DNA sequence (GBSx1604) was identified in *S.agalactiae* <SEQ ID 4655> which encodes the amino acid sequence <SEQ ID 4656>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1934(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4657> which encodes the amino acid sequence <SEQ ID 4658>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1934(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 711/819 (86%), Positives = 767/819 (92%)

Query: 1 MQDKNLVDVNLTSEMKTSTFIDYAMSVIVARALPDVRDGLKPVHRRILYGMNELGVTPDKP 60
MQD+NL+DVNLTSEMKTSTFIDYAMSVIVARALPDVRDGLKPVHRRILYGMNELGVTPDKP
Sbjct: 1 MQDRNLIDVNLTSEMKTSTFIDYAMSVIVARALPDVRDGLKPVHRRILYGMNELGVTPDKP 60

Query: 61 HKKSARITGDVMGKYHPHGDSSIYEAMVRMAQWWSYRHLVDGHGNGFGSMDGDGAAAQRY 120
HKKSARITGDVMGKYHPHGDSSIYEAMVRMAQWWSYRHLVDGHGNGFGSMDGDGAAAQRY
Sbjct: 61 HKKSARITGDVMGKYHPHGDSSIYEAMVRMAQWWSYRHLVDGHGNGFGSMDGDGAAAQRY 120

Query: 121 TEARMSKIALEMLRDINKNTVDFQDNYDGSEREPLVLPARFPNLLVNGATGIAVGMAINI 180
TEARMSKIALE+LRDINKNTV+FDQDNYDGSEREPLVLPARFPNLLVNGATGIAVGMAINI
Sbjct: 121 TEARMSKIALELLRDINKNTVNFQDNYDGSEREPLVLPARFPNLLVNGATGIAVGMAINI 180

Query: 181 PPHNLGESIDAVKLVMDNPDVTTRELMEVIPGDPFPTGALVMGRSGIHRAYETGKGSIVL 240
PPHNL ESIDAVK+VM++PD TTRELMEVIPGDPFPTGALVMGRSGIHRAY+TGKGSIVL
Sbjct: 181 PPHNLAESIDAVKVMHEPDCCTTRELMEVIPGDPFPTGALVMGRSGIHRAYDTGKGSIVL 240

Query: 241 RSRTEIETTSNGKERIVVTEFFPYGVNKTKEVHEHIVRLAQEKRIEGITAVRDESSREGVRF 300
RSRTEIETT G+ERIVVTEFFPYGVNKTKEVHEHIVRLAQEKRL+EGITAVRDESSREGVRF
Sbjct: 241 RSRTEIETTQGRERIVVTEFFPYGVNKTKEVHEHIVRLAQEKRL+EGITAVRDESSREGVRF 300

Query: 301 VIEVRRASANVILNNLFKLTSLQTNFSFNMLAIEKGVPKILSLRQIIDNYIEHQKEVIV 360
VIE+RR ASA VILNNLFKLTSLQTNFSFNMLAIE GVPKILSLRQIIDNYI HQKEVI+
Sbjct: 301 VIEIRREASATVILNNLFKLTSLQTNFSFNMLAIEGVKILSLRQIIDNYISHQKEVII 360

Query: 361 RRTQFDKAKAGARAHILEGLLVALDHLDEVITTIIRNSETDTIAQELMSRFELSERQSQ 420
RRT+FDK KA ARAHILEGLL+ALDHLDEVI IIRNSETD IAQ ELMSRF+LSERQSQ
Sbjct: 361 RRTRPDKAKAEARAHILEGLLIALDHLDEVIAIIRNSETDVIAQELMSRFDLSERQSQ 420

Query: 421 ILDMRLRLTGLERDKIQSEYNDLLALIALDLADILAKPERVVTIIEEMDEVKRYADAR 480
ILDMRLRLTGLERDKIQSEY+DLLALIALDL+DILAKPER++TIIKEEMDE+KRYA+ R
Sbjct: 421 ILDMRLRLTGLERDKIQSEYDDLALIALDLADILAKPERITTIIEEMDEIKRYANPR 480

Query: 481 RTELMI GEVLSLEDEDLIEEEDVLITLSNKGYIKRLAQDEFRAQKRGGRGIQGTGVNND 540
RTELMI+GEVLSLEDEDLIEEEDVLITLSNKGYIKRLAQDEFRAQKRGGRG+QGTGVNND

-1675-

Sbjct: 481 RTEL MVGEVLSLEDEDLIEEEDVLITLSNKG YIKRLAQDEFRAQKRGGRGVQGTGVNND 540

Query: 541 FVREL VSTSTHDTVLFFTNLGRVYRLKAYEIP EYGR TAKGLPIVNLKLDEGETIQTIIIN 600
FVREL+STSTHDT+LFFTN GRVYRLKAYEIP EYGR TAKGLPIVNLKL++GETIQTIIIN

5 Sbjct: 541 FVREL ISTSTHDTLLFFTNFGRVYRLKAYEIP EYGR TAKGLPIVNLKLDEGETIQTIIIN 600

Query: 601 ARKEDVANKYFFFTTQQGIVKRTSVSEFSNIRQNGLRINLKENDELINVLIDENEDVI 660
ARKE+ A K FFFTT+QGIVKRT VSEF+NIRQNGLRA+ LKE D+LINVLL +D+I

10 Sbjct: 601 ARKEETAGKSFFFTTKQGIVKRT EVSEFNIRQNGLRALKLKEGDLINVLITSGQDDII 660

Query: 661 IGTRTGYSVRFKVNAVRNMGRATGVRGVNLREGDKVVGASRIVNGQEVLIITEKGYGKR 720
IGT +GYSVRF ++RNMGR+ATGVRGV LRE D+VVGASRI + QEVL+ITE G+GKR

Sbjct: 661 IGTHSGYSVRFEASIRNMGRSATGVRGVKLREDDRVVGASRIQDNQEVLVITENGFGKR 720

15 Query: 721 TEASEYPTKGRGGKG IKTANITAKNGPLARLVITINGNEDIMVITDTGVIIRTNVANISQT 780
T A++YPTKGRGGKG IKTANIT KNG LA LVT++G EDIMVIT+ GVIIRTNVANISQT

Sbjct: 721 TSATDYPTKGRGGKG IKTANITPKNGQLAGLVTVDGTEDIMVITNKGVIIRTNVANISQT 780

20 Query: 781 GRSTMGVKVMRLDQEA KIVTVALVEQEIEDKSNIEDTKE 819
GR+T+GVK+M+LD +AKIVT LV+ E + I +E

Sbjct: 781 GRATLGVKIMKLDADAKIVTFTLVQPEDSSIAEINTDRE 819

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1515

A DNA sequence (GBSx1605) was identified in *Sagalactiae* <SEQ ID 4659> which encodes the amino acid sequence <SEQ ID 4660>. Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

35

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA04010 GB:AJ000336 L-lactate dehydrogenase [Streptococcus pneumoniae]
Identities = 290/329 (88%), Positives = 313/329 (94%), Gaps = 1/329 (0%)

40 Query: 1 MTATQHKKVILVGDGAVGSSYAFALVNQGI AQELGIIEIPALFDKAVGDAEDLSHALAF 60
MT+TKQHKKVILVGDGAVGSSYAFALVNQGI AQELGIIEIP L +KAVGDA DLSHALAF

Sbjct: 1 MTSTQHKKVILVGDGAVGSSYAFALVNQGI AQELGIIEIPQLHEKAVGDALDLSHALAF 60

45 Query: 61 TSPKKIYAATYADCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVESGFNGI 120
TSPKKIYAA Y+DCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVESGF GI

Sbjct: 61 TSPKKIYAAQYSDCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVESGFNGI 120

50 Query: 121 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALADKIGVDARSVHAYIMGE 180
FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALA+K+ VDARSVHAYIMGE

Sbjct: 121 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALAEKLDVDARSVHAYIMGE 180

55 Query: 181 HGDSEFAVWSHANVAGVQLEQWLQENRDIDEQGLVDLFI SVRDAAYSIIINKKGATYYGIA 240
HGDSEFAVWSHAN+AGV LE++L++ +++ E L++LF VRDAAY+IINKKGATYYGIA

Sbjct: 181 HGDSEFAVWSHANIAGVNLEFLKDTQNVQEAEILIEFEGVRDAAYTIINKKGATYYGIA 240

60 Query: 241 VALARITKAILDDENAVLPLSVYQEGQYGDVVKDFIGQPAIVGAHGIVRPVNIPLNDAEL 300
VALARITKAILDDENAVLPLSV+QEGQYQ V++VFIGQPA+VGAHGIVRPVNIPLNDAE

Sbjct: 241 VALARITKAILDDENAVLPLSVFQEGQYQ-VENVFIGQPAVVGAGHIVRPVNIPLNDAET 299

Query: 301 QKMQASAEQLKDIIDEAWKNPEFQEASKN 329
QKMQASA++L+ IIDEAWKNPEFQEASKN

Sbjct: 300 QKMQASAKELQAIIDEAWKNPEFQEASKN 328

-1676-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4661> which encodes the amino acid sequence <SEQ ID 4662>. Analysis of this protein sequence reveals the following:

5 Possible site: 25

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.17 Transmembrane 106 - 122 (106 - 122)

10 ----- Final Results -----

 bacterial membrane --- Certainty=0.1468(Affirmative) < succ>

 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

15 >GP:AAB81558 GB:U60997 L(+)-lactate dehydrogenase [Streptococcus
 bovis]

 Identities = 278/329 (84%), Positives = 297/329 (89%), Gaps = 2/329 (0%)

20 Query: 1 MTATKQHKVILVGDGAVGSSYAFALVTQNIQELGIIDIFK--EKTQGD AEDLSHALAF 58

 MTATKQHKVILVGDGAVGSSYAFALV Q IAQELGII+I + K GDAEDLSHALAF

 Sbjct: 1 MTATKQHKVILVGDGAVGSSYAFALVNQGIAQELGIIEIPQLFNKAVGDAEDLSHALAF 60

 Query: 59 TSPKKIYAADYSDCHDADLVVITAGAPQKPGETRLDLVEKNLRINKEVVTQIVASGFKGI 118

 TSPKKIYAA Y DC DADLVV+TAGAPQKPGETRLDLV KNL INK +VT++V SGFKGI

25 Sbjct: 61 TSPKKIYAAKYEDCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTEVVKSGFKGI 120

 Query: 119 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALAAKIGVDARSVHAYIMGE 178

 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALA K+ VDARSVHAYIMGE

30 Sbjct: 121 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALAEKLDVDARSVHAYIMGE 180

 Query: 179 HGDSEFAVWSHANVAGVGLYDNLQANRDIDEQGLVDLFI SVRDAAYSIIINKKGATFYGIA 238

 HGDSEFAVWSHANVAGV L +L+ ++++E LV+LF VRDAAYSIIINKKGATFYGIA

 Sbjct: 181 HGDSEFAVWSHANVAGVNLESYLKDVQNVEEAEVLVEFEGVRDAAYSIIINKKGATFYGIA 240

35 Query: 239 VALARITKAILDDENAVLPLSVFQEGQYEGVEDCYIGQPAIVGAYGIVRPVNIPLNDAEL 298

 VALARITKAIL+DENAVLPLSVFQEGQY V DCYIGQPAIVGA+GIVRPVNIPLNDAE

 Sbjct: 241 VALARITKAILNDENAVLPLSVFQEGQYANVTD CYIGQPAIVGAHGIVRPVNIPLNDAEQ 300

40 Query: 299 QKMQASANQLKAIIDEAFAKEEFASAACKN 327

 QKM+ASA +LKAIIDEAF+KEEFASA KN

 Sbjct: 301 QKMEASAKELKAIIDEAFSKEEFASACKN 329

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 286/329 (86%), Positives = 299/329 (89%), Gaps = 2/329 (0%)

 Query: 1 MTATKQHKVILVGDGAVGSSYAFALVNQGIAQELGIIEIPALFDKAVGDAEDLSHALAF 60

 MTATKQHKVILVGDGAVGSSYAFALV Q IAQELGII+I +K GDAEDLSHALAF

 Sbjct: 1 MTATKQHKVILVGDGAVGSSYAFALVTQNIQELGIIDI--FKEKTQGD AEDLSHALAF 58

50 Query: 61 TSPKKIYAATYADCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVEGFGNGI 120

 TSPKKIYAA Y+DC DADLVV+TAGAPQKPGETRLDLV KNL INK +VTQ+V SGF GI

 Sbjct: 59 TSPKKIYAADYSDCHDADLVVITAGAPQKPGETRLDLVEKNLRINKEVVTQIVASGFKGI 118

55 Query: 121 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALADKIGVDARSVHAYIMGE 180

 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALA KIGVDARSVHAYIMGE

 Sbjct: 119 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALAAKIGVDARSVHAYIMGE 178

 Query: 181 HGDSEFAVWSHANVAGVQLEQNLQENRDIDEQGLVDLFI SVRDAAYSIIINKKGATYYGIA 240

 HGDSEFAVWSHANVAGV L WLQ NRDIDEQGLVDLFI SVRDAAYSIIINKKGAT+YGIA

60 Sbjct: 179 HGDSEFAVWSHANVAGVGLYDNLQANRDIDEQGLVDLFI SVRDAAYSIIINKKGATFYGIA 238

 Query: 241 VALARITKAILDDENAVLPLSVYQEGQYGDVKDFIGQPAIVGAHGIVRPVNIPLNDAEL 300

 VALARITKAILDDENAVLPLSV+QEGQY V+D +IGQPAIVGA+GIVRPVNIPLNDAEL

 Sbjct: 239 VALARITKAILDDENAVLPLSVFQEGQYEGVEDCYIGQPAIVGAYGIVRPVNIPLNDAEL 298

-1677-

Query: 301 QKMQASAEQLKDIIDEAWKNPEFQEASKN 329
 QKMQASA QLK IIDEA+ EF A+KN
 Sbjct: 299 QKMQASANQLKAIIDEAFAKEEFASAAKN 327

SEQ ID 4660 (GBS312) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 7; MW 40kDa).

GBS312-His was purified as shown in Figure 205, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1516

A DNA sequence (GBSx1606) was identified in *S.agalactiae* <SEQ ID 4663> which encodes the amino acid sequence <SEQ ID 4664>. This protein is predicted to be NADH oxidase (nox). Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.1888(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC26485 GB:AF014458 NADH oxidase [Streptococcus pneumoniae]
 (ver 2)
 Identities = 363/458 (79%), Positives = 408/458 (88%), Gaps = 3/458 (0%)

Query: 1 MSKIVVVGTHAGTAAIKTMSNYGEANEIVTFDQNSNISFLGCGMALWIGEQIDGPEGL 60
 MSKIVVVG NHAGTA I TML N+G NEIV FDQNSNISFLGCGMALWIGEQIDG EGL
 Sbjct: 1 MSKIVVVGANHAGTACINTMLDNFGNENEIVVFDQNSNISFLGCGMALWIGEQIDGA EGL 60

Query: 61 FYSDKEQLESMAKAVYMNSPVLNIDYDKKEVTALVDGKEHVESYEKILILATGSQPIIPPI 120
 FYSDKE+LE+ GAKVYMNSPVL+IDYD K VTA V+GKEH ESYEKLI ATGS PI+PPI
 Sbjct: 61 FYSDKEKLEAKGAKVYMNSPVLNIDYDNKVVTAEEVGEKHEKESYEKILIFATGSTPILPPI 120

Query: 121 KGVEIQEGSREFKATLENLQFVKLYQNSEEVIEKLAKEG--INRVAVVGAGYIGVELAEA 178
 +GVEI +G+REFKATLEN+QFVKLYQN+EEVI KL+ ++R+AVVG GYIGVELAEA
 Sbjct: 121 EGVEIVKGNREFKATLENVQFVKLYQNAEEVINKLSDKSQHLDRIVVGGGYIGVELAEA 180

Query: 179 FQRIGKEVTLVDVADTCMGYYDRDFTDMMSKNLEDHGI RLAFGQAVQAVEGDGKVERLV 238
 F+R+GKEV LVD+ DT + GYYD+DFT MM+KNLEDH IRLA GQ V+A+EGDGKVERL+
 Sbjct: 181 FERLGKEVVLVDIVDTVLNGYYDKDFTQMMAKNLEDHNIRLALGQTVKAIEGDGKVERLI 240

Query: 239 TDKETFDVDMVILAVGFRPNTLGAAGKLDTPFRNGAWVVDKKQETSVDVYAIGDCATIWD 298
 TDKE+FDVDMVILAVGFRPNT L GK++ FRNGA++VDKKQETS+ VYA+GDCAT++D
 Sbjct: 241 TDKESFDVDMVILAVGFRPNTALADGKIELFRNGAFLVDKKQETSIPGVYAVGDCATVVD 300

Query: 299 NSRDDINIALASNAVRTGIVAAHNACGTELEGAGVQGSNGISYGLNMVSTGLTLEKAK 358
 N+R D +YIALASNAVRTGIV A+NACG ELEGG VQGSNGISYGL+MVSTGLTLEKAK
 Sbjct: 301 NARKDTSYIALASNAVRTGIVGAYNACGHELEGIGVQGSNGISYGLHVMVSTGLTLEKAK 360

Query: 359 QAGYNAVETGFNDLQKPEFIKHNNEHVAIKIVYDKDSRVILGCQMVSHD-DVSMGIHMF 417
 AGYNA ETGFNDLQKPEF+KH+NHEVAIKIV+DKDSR ILG QMVSH+ +SMGIHMF
 Sbjct: 361 AAGYNATETGFNDLQKPEFMKHDNHEVAIKIVFDKDSREILGQMVSHDIAISMGIHMF 420

Query: 418 LAIQEKVTIEKLALTDIFFLPHFNKPYNYITMAALGAK 455
 LAIQE VTI+KLALTD+FFLPHFNKPYNYITMAAL A+
 Sbjct: 421 LAIQEHVTIDKLALTDLFFLPHFNKPYNYITMAALTAE 458

-1678-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4665> which encodes the amino acid sequence <SEQ ID 4666>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2068(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 362/456 (79%), Positives = 403/456 (87%)

15 Query: 1 MSKIVVVG TNHAGTAAIKTMLS NYGEANEIVTFDQNSNISFLGCGMALWIGE QIDGPEGL 60
 Sbjct: 1 MSKIVVVG NHAGTAAIKTMLS NYGEANEIVFDQNSNISFLGCGMALWIGE QI GPEGL 60

20 Query: 61 FYSDKEQLES MGAKVYMNSFVLNIDYDKKEVTALVDGKEHVESYEKLILATGSQPIIPPI 120
 Sbjct: 61 FYSDKE+LES+GAKVYM SPV +IDYD K VTALVDGK HVE+Y+KLI ATGSQPI+PPI 120

25 Query: 121 KGVEIQEGSREFKATLENLQFVKLYQNSSEEVIEKLAKPGINRVAVVGAGYIGVELAEAFQ 180
 Sbjct: 121 KG EI+EGS EF+ATLENLQFVKLYQNS +VI KL I RVAVVGAGYIGVELAEAFQ 180

30 Query: 181 RIGKEVTLVDVADTCMGYYDRDFTDMMSKNLEDHGIRLAFGQAVQAVEGDGKVERLVID 240
 Sbjct: 181 RKGKEVVLIDVVDTCLAGYYDRDLTDLMAKNMEEHGIQLAFGETVKEVAGNGKVEKIITD 240

35 Query: 241 KETFDVDMVILAVGFRPNTLGGAGKLDTFRNGAWVVDKKQETSVKDVYAIGDCATIWDNS 300
 Sbjct: 241 KNEYDVMVILAVGFRPNTTLGNGKIDLFRNGAFLVNRQETSIPGVYAIGDCATTYDNA 300

40 Query: 301 RDDINYIALASNAVRTGIVA AHNACGTELEGAGVQGSNGISITGLNMVSTGLTLEKAKQA 360
 Sbjct: 301 TRDTNYIALASNAVRTGIVA AHNACGTDLEGIGVQGSNGISITGLHNMVSTGLTLEKAKRL 360

45 Query: 361 GYNAVETGFNDLQKPEFIKHNNHEVAIKIVYDKDSRVILGCQMVSHEDVSMGIHMFSLAI 420
 Sbjct: 361 GFDAAVTEYTDNQKPEFIKHGNNFPVTIKIVYDKDSRRILGAQMAAREDVSMGIHMFSLAI 420

Query: 421 QEKVTIEKLALTDIFFLPHFNKPNYITMAALGAKD 456
 Sbjct: 421 QEGVTIEKLALTDIFFLPHFNKPNYITMAALGAKD 456

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1517

50 A DNA sequence (GBSx1607) was identified in *S.agalactiae* <SEQ ID 4667> which encodes the amino acid sequence <SEQ ID 4668>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2319(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

60 The protein has no significant homology with any sequences in the GENPEPT database.

-1679-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1518

- 5 A DNA sequence (GBSx1608) was identified in *S.agalactiae* <SEQ ID 4669> which encodes the amino acid sequence <SEQ ID 4670>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

10	INTEGRAL	Likelihood = -7.75	Transmembrane	160 - 176 (157 - 179)
	INTEGRAL	Likelihood = -7.38	Transmembrane	73 - 89 (70 - 97)
	INTEGRAL	Likelihood = -5.47	Transmembrane	289 - 305 (284 - 312)
	INTEGRAL	Likelihood = -4.09	Transmembrane	107 - 123 (106 - 124)
	INTEGRAL	Likelihood = -3.24	Transmembrane	43 - 59 (43 - 59)
	INTEGRAL	Likelihood = -1.91	Transmembrane	258 - 274 (258 - 275)
15	INTEGRAL	Likelihood = -1.33	Transmembrane	234 - 250 (233 - 251)
	INTEGRAL	Likelihood = -0.00	Transmembrane	209 - 225 (209 - 225)

----- Final Results -----

20	bacterial membrane --- Certainty=0.4100 (Affirmative) < succ>
	bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
	bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9805> which encodes amino acid sequence <SEQ ID 9806> was also identified.

- 25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15146 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
Identities = 172/318 (54%), Positives = 234/318 (73%)

30	Query: 5	LSLTTIFALLFSSMLIYATPLIFTSIGGTFSEKGGIVNVGLEGIMVIGAFSGVVFNFLEFA	64
		+ + I + + + L+YA PLI T++GG FSER G+VN+GLEG+M+IGAF+ V+PNL F	
	Sbjct: 1	MDIVQILSLIIVPATLVYAAPLILITLALGGVFSERSGVNVNIGLEGLMIIGAFTSVLFNLFPG	60
35	Query: 65	SVFGDAPWISVLVGGLVGLIFSVIHAVATVNFADHIIISGTVLNLMAPSLAVFLIKVLY	124
		G A PW+S+L G +FS+IHA A ++FRAD +SG +N++A +F++K++Y	
	Sbjct: 61	QELGAAPWLSLLAAMAAGALFSLIHAAAISFRADQTVSGVAINMLALGATLFIVKLIY	120
40	Query: 125	NKGQTDNIQESFGKFNFPILSDIPFVGDIFFKGTSLVGYYIALIFSFLAWFILYKTRFGLR	184
		K QTD I E F K P L DIP +G IFF +AI +F++WFIL+KT FGLR	
	Sbjct: 121	GKAQTDKIPEPFYKTKIPGLGDIPVLGKIFFSDVYYSILALALAFISWILFKTPFGLR	180
45	Query: 185	LRSVGEHPQAADTLGINVYLMRYSGVLISGFLGGIGGAVYAQSISVNFAATTILGPGFIS	244
		+RSVGEHP AADT+GINVY MRY GV+ISG GG+GG VYA +I+++F +TI G GFI+	
	Sbjct: 181	IRSVGEHPMAADTMGINVYKMRIGVMISGLFGLGGVYASTIALDFTHSTISGQGFIA	240
50	Query: 245	LAAMIFGKWNPIGAMLASLFFGLSQSLAVIGSHLPLLSNIPTVYLIAPYVLTIIIVLAAF	304
		LAA++FGKW+PIGA+ A+LFFG +QSL++IGS LPL +IP VY+ +APY+LTI+ L F	
	Sbjct: 241	LAALVFGKWHPIGALGAALFFGFAQSLSIIGSLPLPKDIPNVYMLMAPYILTILALTGF	300
	Query: 305	FGQAVAPKADGINYIKTK	322
		G+A APKA+G+ YIK K	
	Sbjct: 301	IGRADAPKANGVPYIKGK	318

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4671> which encodes the amino acid sequence <SEQ ID 4672>. Analysis of this protein sequence reveals the following:

55 Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -8.92	Transmembrane	73 - 89 (69 - 97)
----------	--------------------	---------------	--------------------

-1680-

INTEGRAL Likelihood = -5.04 Transmembrane 160 - 176 (158 - 177)
 INTEGRAL Likelihood = -4.62 Transmembrane 289 - 305 (284 - 312)
 INTEGRAL Likelihood = -3.98 Transmembrane 234 - 250 (232 - 251)
 INTEGRAL Likelihood = -2.13 Transmembrane 107 - 123 (106 - 123)
 5 INTEGRAL Likelihood = -2.02 Transmembrane 43 - 59 (43 - 59)
 INTEGRAL Likelihood = -0.53 Transmembrane 258 - 274 (258 - 274)

----- Final Results -----
 10 bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

15 >GP:CAB15146 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 176/318 (55%), Positives = 239/318 (74%)

Query: 5 MSLVTIFALLMSSMLIYATPLIFTSIGGTFSESGVNVGLEGIMVMGAFSGIVFNLEFA 64
 M +V I +++ + L+YA PLI T++GG FSERSGVVN+GLEG+M++GAF+ ++FNL F
 20 Sbjct: 1 MDIVQILSIIVPATLVYAAPLILITALGGVFSESGVNVNIGLEGLMIGAFSTVLFNLF 60

Query: 65 ETFGKATPWIAVLVGGIVGLIFSLIHAVATINFRADHIVSGTVLNLAPSFAVFLVKAMY 124
 + G A PW+++L G +FSLIHA A I+FRAD VSG +N+LA +F+VK +Y
 25 Sbjct: 61 QELGAAAPWLSLLAAMAAGALFSLIHAAAISFRADQTVSGVAINMLALGATLFIVKLIY 120

Query: 125 GKGQTDNIQQSFGKFDFFPGLSQIPVIGDIFFKNTSLIGYFAIAFSFFAWFLLYKTRFGLR 184
 GK QTD I + F K PGL IPV+G IFF + AIA +F +WF+L+KT FGLR
 30 Sbjct: 121 GKAQTDKIPEPFYKTKIPGLGDIPVLGKIFFSDVYYSILALALAFISWIFLKTTPFGLR 180

Query: 185 LRSVGEHPQAADTLGINVYLMKYVGVMISGFLGGIGGAVYAQSISVNFVTTILGPGFIA 244
 +RSVGEHP AADT+GINVY M+Y GVMISG GG+GG VYA +I+++F +TI G GFIA
 35 Sbjct: 181 LRSVGEHPMAADTMGINVYKMYIGVMISGLFGLGGGVYASTIALDFTHSTISGQGFIA 240

Query: 245 LAAMIFGKWNPGAMLSLFFGLSQSLAVIGAQLPLLEKIPTVYLQIAPYMTIILAAF 304
 LAA++FGKW+P+GA+ ++LFFG +QSL++IG+ LPL + IP VY+ +APY++TI+ L F
 40 Sbjct: 241 LAALVFGKWHPIGALGAALFFGFAQSLSIIGSLPLPKDIPNVYMLMAPYILTILALTGF 300

Query: 305 FGQAVAPKADGINYIKSK 322
 G+A APKA+G+ YIK K
 45 Sbjct: 301 IGRADAPKANGVPYIKGK 318

An alignment of the GAS and GBS proteins is shown below.

Identities = 272/322 (84%), Positives = 301/322 (93%)

45 Query: 1 MVSKLSTTIFALLFSSMLIYATPLIFTSIGGTFSESGVNVGLEGIMVIGAFSGVVFN 60
 +V+K+SL TIFALL SSMLIYATPLIFTSIGGTFSESGVNVGLEGIMV+GAFSG+VFN
 Sbjct: 1 VVNKMSLVITIFALLMSSMLIYATPLIFTSIGGTFSESGVNVGLEGIMVMGAFSGIVFN 60

Query: 61 LEFASVFGDATPWISVLVGGVLGLIFSVIHAVATVNFADHIIISGTVLNLMAPSLAVFLI 120
 LEFA FG ATPWI+VLVGG+VGLIFS+IHAVAT+NFRADHI+SGTVLNL+APS AVFL+
 50 Sbjct: 61 LEFAETFGKATPWIAVLVGGIVGLIFSLIHAVATINFRADHIVSGTVLNLAPSFAVFLV 120

Query: 121 KVLNKGQTDNIQESFGKFNFPILSDIPFVGDIFFKNTSLVGYIAILFSFLAWFILIYKTR 180
 K +Y KGQTDNIQ+SFGKF+FP LS IP +GDIFFK TSL+GY AI FSF AWF+LYKTR
 55 Sbjct: 121 KAMYKGQTDNIQQSFGKFDFFPGLSQIPVIGDIFFKNTSLIGYFAIAFSFFAWFLLYKTR 180

Query: 181 FGLRLRSVGEHPQAADTLGINVYLMRYSGVLSGFLGGIGGAVYAQSISVNFATTTILGP 240
 FGLRLRSVGEHPQAADTLGINVYLM+Y GV+ISGFLGGIGGAVYAQSISVNFATTTILGP
 60 Sbjct: 181 FGLRLRSVGEHPQAADTLGINVYLMKYVGVMISGFLGGIGGAVYAQSISVNFATTTILGP 240

Query: 241 GFISLAAMIFGKWNPGAMLASLFFGLSQSLAVIGSHLPLLSNIPTVYLQIAPYVLTIIIV 300
 GFI+LAAMIFGKWNPGAML+SLFFGLSQSLAVIG+ LPLL IPTVYLQIAPY++TII+
 65 Sbjct: 241 GFIALAAMIFGKWNPGAMLSLFFGLSQSLAVIGAQLPLLEKIPTVYLQIAPYMTIIII 300

Query: 301 LAAFFGQAVAPKADGINYIKTK 322
 LAAFFGQAVAPKADGINYIK+K
 70 Sbjct: 301 LAAFFGQAVAPKADGINYIKSK 322

-1681-

A related GBS gene <SEQ ID 8829> and protein <SEQ ID 8830> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1      Crend: 3
McG: Discrim Score:      8.61
GvH: Signal Score (-7.5): -1.53
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
ALOM program      count: 8 value: -7.75 threshold: 0.0
10  INTEGRAL      Likelihood = -7.75      Transmembrane 160 - 176 ( 157 - 179)
    INTEGRAL      Likelihood = -7.38      Transmembrane 73 - 89 ( 70 - 97)
    INTEGRAL      Likelihood = -5.47      Transmembrane 289 - 305 ( 284 - 312)
    INTEGRAL      Likelihood = -4.09      Transmembrane 107 - 123 ( 106 - 124)
15  INTEGRAL      Likelihood = -3.24      Transmembrane 43 - 59 ( 43 - 59)
    INTEGRAL      Likelihood = -1.91      Transmembrane 258 - 274 ( 258 - 275)
    INTEGRAL      Likelihood = -1.33      Transmembrane 234 - 250 ( 233 - 251)
    INTEGRAL      Likelihood = -0.00      Transmembrane 209 - 225 ( 209 - 225)
    PERIPHERAL    Likelihood = 3.34      139
modified ALOM score: 2.05

20  *** Reasoning Step: 3

----- Final Results -----
        bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

30  ORF00914(313 - 1266 of 1566)
    EGAD|108729|BS3151(1 - 318 of 319) hypothetical protein {Bacillus
    GP|1934814|emb|CAB07939.1||Z93937 unknown {Bacillus
    GP|2635653|emb|CAB15146.1||Z99120 similar to hypothetical proteins {Bacillus
    PIR|F70009|F70009 conserved hypothetical protein yufQ - Bacillus subtilis
    %Match = 34.9
35  %Identity = 54.1 %Similarity = 76.4
    Matches = 172 Mismatches = 75 Conservative Sub.s = 71

    216      246      276      306      336      366      396      426
40  TLQVFHLS*LKL*QLQSSSS*VSITLTL SMLNLNKNK*KVVS KLSLTTIFALLFSSMLIYATPLIFTSIGGTF SERGGIVN
    : : |::: : :| | |::| | | | | | | |
    MDIVQILSIIVPATLVYAAPLILALTALGGVFSERSGVVN
    10      20      30

    456      486      516      546      576      606      636      666
45  VGLEGIMVIGAFSGGVFNLEFASVFGDATPFWISVLVGGVLGLIFSVIHAVATVNF RADHIISGTVLNLMAPSLAVFLIKV
    : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
    IGLEGLMIIGAFTSVLFNLFQQLGAAAPWLSLLAAMAAGALFSLIHAAAISFRADQTVSGVAINMLALGATLFIIVKL
    50      60      70      80      90      100      110

    696      726      756      786      816      846      876      906
50  LYNKQGTNDNIQESFGKFNFPILSDIPFVGDIFFKGTSLVGVIALLFSFLAWFILYKTRFGLRLRSVGEHPQAADTLGINV
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    IYGKAQTDKIPEPFYKTKIPGLGDIPVLGKIFFSVDVYTSILAIALAFISWFILFKTPFGLRIRSVGEHPMAADTMGINV
    130      140      150      160      170      180      190

    936      966      996      1026      1056      1086      1116      1146
55  YLMRYSGVLISGLGGIGGAVYAQSISVNF AATTILGPGFISLAAMIFGKWNPIGAMLASLFPGLSQSLAVIGSHLPLLS
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    YKMYRIGVMISGLFGGLGGGVYASTIALDFTHSTISGQGFIALAALVFGKWHPIGALGAALFFGFAQSLSIIGSLPLPFK
60  210      220      230      240      250      260      270

    1176      1206      1236      1266      1296      1326      1356      1386
NIPTVYLQIAPYVLTIIIVLAFFGQAVAPKADGINYIKTK*IKRN*YKLVSFYCL*ICEKILCENFT*IIIQ*Q*NIKK*
    : | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    DIPNVYMLMAPYILLTILALTGFIGRADAPKANGVPIYIKGR

```

-1682-

290 300 310

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1519

A DNA sequence (GBSx1609) was identified in *S.agalactiae* <SEQ ID 4673> which encodes the amino acid sequence <SEQ ID 4674>. This protein is predicted to be ribose/galactose ABC transporter, permease protein (rbsC-1). Analysis of this protein sequence reveals the following:

```

Possible site: 55
10 >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL Likelihood = -14.59 Transmembrane 205 - 221 ( 200 - 228)
    INTEGRAL Likelihood = -13.69 Transmembrane 21 - 37 ( 13 - 45)
    INTEGRAL Likelihood = -7.27 Transmembrane 302 - 318 ( 290 - 321)
    INTEGRAL Likelihood = -7.17 Transmembrane 115 - 131 ( 111 - 138)
15 INTEGRAL Likelihood = -4.25 Transmembrane 251 - 267 ( 250 - 268)
    INTEGRAL Likelihood = -2.97 Transmembrane 63 - 79 ( 63 - 80)
    INTEGRAL Likelihood = -2.87 Transmembrane 333 - 349 ( 328 - 349)

----- Final Results -----
20 bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8831> which encodes amino acid sequence <SEQ ID 8832> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1 Crend: 6
SRCLG: 0
McG: Length of UR: 24
    Peak Value of UR: 3.06
30 Net Charge of CR: 3
McG: Discrim Score: 12.53
GvH: Signal Score (-7.5): -5.31
    Possible site: 46
    >>> Seems to have an uncleavable N-term signal seq
35 Amino Acid Composition: calculated from 1
    ALOM program count: 7 value: -14.59 threshold: 0.0
    INTEGRAL Likelihood = -14.59 Transmembrane 196 - 212 ( 191 - 219)
    INTEGRAL Likelihood = -13.69 Transmembrane 12 - 28 ( 4 - 36)
    INTEGRAL Likelihood = -7.27 Transmembrane 293 - 309 ( 281 - 312)
40 INTEGRAL Likelihood = -7.17 Transmembrane 106 - 122 ( 102 - 129)
    INTEGRAL Likelihood = -4.25 Transmembrane 242 - 258 ( 241 - 259)
    INTEGRAL Likelihood = -2.97 Transmembrane 54 - 70 ( 54 - 71)
    INTEGRAL Likelihood = -2.87 Transmembrane 324 - 340 ( 319 - 340)
    PERIPHERAL Likelihood = 0.16 133
45 modified ALOM score: 3.42
icml HYPID: 7 CFP: 0.684

*** Reasoning Step: 3

50 ----- Final Results -----
    bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

55 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB15145 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
Identities = 154/349 (44%), Positives = 220/349 (62%), Gaps = 6/349 (1%)

```

```

Query: 10 MSKKAQKIADVPLISVVLGIIILGAIIMLIFGYDPLWGYEGLFQTAFSGSIKNIGEIFRAMGP 69

```

-1683-

M K+ + VPLI+++LG+ GA+IML+ GY GY L+ FG I +GE R + P
 Sbjct: 1 MVKRLSHLLVPLIAIILGLAAGALIMLVSGYSVASGYSALWNGIFGEIYYVGETIRQITP 60
 Query: 70 LILIALGFSVASRAGFFNIGLPGQALSGWIAAGWFALSHPDMPRPAMILCTIIIGIVAGG 129
 5 IL L + A R G FNIG+ GQ L GW AA W + D P + +I AGG
 Sbjct: 61 YILSGLAVAFARTGLFNIGVEGQLLVGWTAAVVVGTAFA-DGPAYIHLPLALITAAAAGG 119
 Query: 130 ITGAIPGILRAYLGTSEVIVTIMMNYIVLYSGNAIVQRFVFPKSIMRTSDSSVYVSANASY 189
 + G IPGIL+A EVIVTIMMNYI L+ N I+ V D + + +AS
 10 Sbjct: 120 LWGFIPGILKARFYVHEVIVTIMMNYIALHMTNYIISNVLTDH----QDKTGKIHESASL 175
 Query: 190 QTDWLSLSTNNSRINIGIFIAIIAVVLVWFLNKTTLGFEIRSVGLNPNASEYAGMSAKR 249
 ++ +L +T+ SR+++GI +A++A V++WF++NK+T GFE+R+VG N +AS+YAGMS ++
 15 Sbjct: 176 RSPFLEQITDYSRLHLGLIIVALLAAVIMWFIINKSTKGFEIRAVGFNQHASQYAGMSVRK 235
 Query: 250 TIILSMIISGAFAGLGGVVEGLGTFFENVFQPSLAIGFDGMAVSLAANSPIGILFAAF 309
 I+ SM+ISGAFAGL G +EGLGTFE V+ + +GFDG+AV+LL N+ +G++ AA
 Sbjct: 236 NIMTSMIISGAFAGLAGAMEGLGTFFEYAAVKGAFTGVGFDGIAVALGGNTAVGVVLAAC 295
 20 Query: 310 LFGVLSVGAPGMNI-AGIPPELIKVVVTASIIFFVGVHYIIEYVIKPKKQ 357
 L G L +GA M I +G+P E++ +V A II FV Y I +V+ K+
 Sbjct: 296 LLGGLKIGALNMPIESGVPSEVVDIVIAIILFVASSYAIRFVMGKLKK 344

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2149> which encodes the amino acid
 25 sequence <SEQ ID 2150>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq
 30 INTEGRAL Likelihood = -12.74 Transmembrane 205 - 221 (200 - 228)
 INTEGRAL Likelihood = -12.42 Transmembrane 21 - 37 (14 - 45)
 INTEGRAL Likelihood = -7.22 Transmembrane 115 - 131 (111 - 135)
 INTEGRAL Likelihood = -4.78 Transmembrane 251 - 267 (249 - 269)
 INTEGRAL Likelihood = -2.50 Transmembrane 70 - 86 (69 - 86)
 INTEGRAL Likelihood = -2.34 Transmembrane 302 - 318 (300 - 318)
 35 INTEGRAL Likelihood = -1.44 Transmembrane 148 - 164 (147 - 165)
 INTEGRAL Likelihood = -1.33 Transmembrane 326 - 342 (326 - 342)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 293/358 (81%), Positives = 333/358 (92%), Gaps = 1/358 (0%)
 45 Query: 6 RRREMSKKAQKIAVPLISVVLGIIILGAIIMLIFGYDPLWGYEGLFQTAFGSIKNIGEIPR 65
 RR+ MSK AQKIAVPLISV+LG +LGAIIM+IFGYDP+WGYEGLFQ AFGS+KNIGEIPR
 Sbjct: 6 RRVMSKNAQKIAVPLISVVLGIIILGAIIMVIFGYDPIWGYEGLFQIAFGSVKNIGEIPR 65
 Query: 66 AMGPLILIALGFSVASRAGFFNIGLPGQALSGWIAAGWFALSHPDMPRPAMILCTIIIGI 125
 50 +MGPLILIALGF+VASRAGFFN+GL GQAL+GWI+AGWFAL +PDMRP +IL T +IG+
 Sbjct: 66 SMGPLILIALGFTVASRAGFFNVGLSGQALAGWISAGWFALNPDMPRLILMTALIGM 125
 Query: 126 VAGGITGAIPGILRAYLGTSEVIVTIMMNYIVLYSGNAIVQRFVFPKSIMRTSDSSVYVSA 185
 +AGGI GAIPGILRAYLGTSEVIVTIMMNYI+LY GNAIVQR +P+S+ ++ DS++ VS
 55 Sbjct: 126 IAGGIAGAIPGILRAYLGTSEVIVTIMMNYIILYVGNNAIVQRGYPESVKQSIDSTIQVSD 185
 Query: 186 NASYQTDWLSLSTNNSRINIGIFIAIIAVVLVWFLNKTTLGFEIRSVGLNPNASEYAGM 245
 NASYQT WLS+LTNNSRINIGIF AIIA+ L+WFLNKTTLGFEIRSVGLNP+ASEYAGM
 60 Sbjct: 186 NASYQTHWSALTNNSRINIGIFFAIIAIALIWFLLNKTTLGFEIRSVGLNPNASEYAGM 245
 Query: 246 SAKRTIILSMIISGAFAGLGGVVEGLGTFFENVFQPSLAIGFDGMAVSLAANSPIGIL 305
 S+KRTIILSMIISGA AGLGGVVEGLGTFFENVFQ SSLA+GFDGMAVSLAANSPIGIL
 Sbjct: 246 SSKRTIILSMIISGALAGLGGVVEGLGTFFENVFQSSSLAVGFDGMAVSLAANSPIGLIF 305
 65 Query: 306 FAAFLEGVLSVGAPGMNIAGIPPELIKVVVTASIIFFVGVHYIIE-YVIKPKKQMKGGK 362

-1684-

F++FLFGVL++GAPGMNIAGIPPEL+KVV+ASIIFFVG HY+IE Y+I+PKK +KGGK
 Sbjct: 306 FSSFLFGVLNIGAPGMNIAGIPPELVKVV+ASIIFFVGSHYLIERYIIRPKKLVKGGK 363

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1520

A DNA sequence (GBSx1610) was identified in *S.agalactiae* <SEQ ID 4675> which encodes the amino acid sequence <SEQ ID 4676>. This protein is predicted to be sugar ABC transporter, ATP-binding protein (mglA). Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3851(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9803> which encodes amino acid sequence <SEQ ID 9804> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15144 GB:Z99120 similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]
 Identities = 311/497 (62%), Positives = 396/497 (79%), Gaps = 1/497 (0%)

Query: 14 VIEMKEITKKFGDFVANDHINLTVEKGEIHALLGENGAGKSTLMNMLAGLLEPTDGGQIFI 73
 VIEM I K F VAND+INL V+KGEIHALLGENGAGKSTLMN+L GL +P G+I +
 Sbjct: 4 VIEMLNIRKAPPGIVANDNINLQVKKGEIHALLGENGAGKSTLMNVLFGLYQPERGEIRV 63

Query: 74 NGQPVTIDSPSKSSQLGIGMVHQHFMLVEAFTVAENIVLGNETTQNGVLDIKTAAKEIKE 133
 G+ V I+SP+K++ LGIGMVHQHFMLV+ FTVAENI+LG E + G +D K A +E+++
 Sbjct: 64 RGEKVHINSPNKANDLGIGMVHQHFMLVDTFTVAENIILGKEPKKFGRIDRKRAQGEVQD 123

Query: 134 LSEKYGLSVNPNNAKISDISVGAQQRVEILKTLRGADILIFDEPTAVLTPSEIKELMTIM 193
 +S++YGL ++P AK +DISVG QQR EILKTLRGADILIFDEPTAVLTP EIKELM IM
 Sbjct: 124 ISDRYGLQIHPEAKAADISVGMQQRRAEILKTLRGADILIFDEPTAVLTPHEIKELMQIM 183

Query: 194 KSLVKEGKSIILITHKLDEIRAVADKVTVIRRGKSIETVPVAGASSQQLAEMMVGRSVSF 253
 K+LVKEGKSIILITHKL EI + D+VTVIR+GK I+T+ V + +LA +MVGR VSF
 Sbjct: 184 KNLVKEGKSIILITHKLKEIMEICDRVTVIRKKGKIKTLDVDRDTNQDELASLMVGREVSF 243

Query: 254 RTEKKEANPTDIILSVKDLVVEENRGGVLAVKNLSLDVRAGEIVGIAGIDGNGQSELIQA 313
 +TEK+ A P +L++ + V++ R G+ V++LSL V+AGEIVGIAG+DGNGQSELI+A
 Sbjct: 244 KTEKRAAQPGAELVLAIDGITVKDTR-GIETVRDL SLSVKAGEIVGIAGVDGNGQSELI EA 302

Query: 314 ITGLRKVTSGQIVIKGKDVTKFSSRQITELSVGHVPEDRHRDGLVLDMTAENLALQTY 373
 +TGLRK SG I + GK + + R+ITE +GH+P+DRH+ GLVLD + EN+ LQ+YY
 Sbjct: 303 VTGLRKTDSTGTTITLNGKQIQNLTPRKITESGIGHIPQDRHKHGLVLDFFIGENILLQSY 362

Query: 374 KEPLSHKGILNFAKIEYARQLMTEFDVRGAGEHVLARGFSGGNQKAIAREVDRDPDL 433
 K+P S G+L+ ++ + AR L+TE+DVR E+ AR SGGNQKAI RE+DR+PDL
 Sbjct: 363 KKPYSALGVLHKGEMYKARSLLITEYDVRTPDEYTHARALSGGNQKALIGREIDRNPDL 422

Query: 434 LIVSQPTRGLDVGAIEYIHKRLIEERDKGKAVLVVSPFELDEILNLSDRIAVIHDGKIQGI 493
 LI +QPTRGLDVGAIE++HK+LIE+RD GKAVL++SPFEL+EI+NLSDRIAVI +G+I
 Sbjct: 423 LIAAQPTRGLDVGAIEFVHKRLIEQRDAGKAVLLLSFELEEIMNLSDRIAVIFEGRIIAS 482

Query: 494 VKPDQTNKQELGILMAG 510
 V P +T +QELG+LMAG

-1685-

Sbjct: 483 VNPQETTEQELGLLMAG 499

Identities = 75/242 (30%), Positives = 128/242 (51%), Gaps = 24/242 (9%)

5

Query: 280 GVLAVKNLSLDVRAGEIVGIAGIDGNGQSELIQAITGLRKVTSGQIVIKGKDVTKFSSRQ 339

G++A N++L V+ GEI + G +G G+S L+ + GL + G+I ++G+ V S +

Sbjct: 16 GIVANDNINLQVKKGEIHALLGENGAGKSTLMNVLFGLYQPERGEIRVRGEKVVHINSPNK 75

10

Query: 340 ITEL SVGHVPEDRHRDGLVLD-MTMAENLALQTYTKEPLSHKGILNFAKI--KEYARQLM 396

+L +G V H+ +++D T+AEN+ L KEP F +I K +++

Sbjct: 76 ANDLGIGMV---HQHFMLVDTFTVAENIILG---KEPKK-----FGRIDRKRAQGEVQ 122

15

Query: 397 TEF DVRGAGEHVLARG--FSGGNQQAIIAREVDRDPDLLIVSQPTRGL---DVGAI EYI 451

D G H A+ S G QQ+A I + + R D+LI +PT L ++ + I

Sbjct: 123 DISDRYGLQIHPEAKAADISVGMQRAEILKTL YRGADILIFDEPTAVLTPHEIKELMQI 182

Query: 452 HKRLIEERDKGKAVLVVSFELDEILNLSDRIAVIHDGKIQQGIVKPDQTNKQELGILMAGG 511

K L++E GK+++++ +L EI+ + DR+ VI GK + TN+ EL LM G

Sbjct: 183 MKNLVKE---GKSIILITHKLKEIMEICDRVTVIRKGGKGIKTL DVRTNQDELASLMVGR 239

20

Query: 512 KI 513

++

Sbjct: 240 EV 241

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4677> which encodes the amino acid
 25 sequence <SEQ ID 4678>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

30

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3558(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35

Identities = 431/511 (84%), Positives = 467/511 (91%), Gaps = 1/511 (0%)

Query: 10 MTQNVIMKEITKKFGDFVANDHINLTVEKGEIHALLGENGAGKSTLMNMLAGLLEPTDG 69

MTQ+VIEM+BITKKFGDFVANDHINL V KGEIHALLGENGAGKSTLMNMLAGLLEPT G

40

Sbjct: 7 MTQNVIMREITKKFGDFVANDHINLNVKGEIHALLGENGAGKSTLMNMLAGLLEPTSG 66

Query: 70 QIFINGQPVITIDSPSKSSQLGIGMVHQHFMLVEAFTVAENIVLGNETTQNGVLDIKTAAK 129

+I IN +PV IDSPSKS++LGIGMVHQHFMLVEAFTVAENI+LGNE +NG LD+ A+K

Sbjct: 67 BIVINDKPVQIDSPSKSAKLIGIGMVHQHFMLVEAFTVAENIILGNEVVKNGCLDLNQASK 126

45

Query: 130 EIKELSEKYGLSVNPNAKISDISVGAQQRVEILKTL YRGADILIFDEPTAVLTPSEIKEL 189

+IK LSEKYGL++NP+AK+SDISVGAQQRVEILKTL YRGADILIFDEPTAVLTP+EIKEL

Sbjct: 127 DIKVLSEKYGLAINPSAKVSDISVGAQQRVEILKTL YRGADILIFDEPTAVLTPAEIKEL 186

50

Query: 190 MTIMKSLVKEGKSIILITHKLDEIRAVADKVTVIRRGKSIETVPVAGASSQQLAEMMVGR 249

MTIMK+LVKEGKSIILITHKLDEIRAVAD+VTVIRRGKSIETV VAGA+SQ LAEMMVGR

Sbjct: 187 MTIMKNLVKEGKSIILITHKLDEIRAVADRVTVIRRGKSIETVDVAGATSDLAEMMVGR 246

55

Query: 250 SVSFRTEKKEANPTDIILSVKDLVVEENRGGVLAVKNLSLDVRAGEIVGIAGIDGNGQSE 309

SVSF T KK A P D++LS+K+L V+ENR GV AVK LSLDVRAGEIVGIAGIDGNGQSE

Sbjct: 247 SVSFTTSKKAABPKDVVLSIKNLEVDENR-GVPAVKGLSLDVRAGEIVGIAGIDGNGQSE 305

60

Query: 310 LIQAITGLRKVTSGQIVIKGKDVTKFSSRQITEL SVGHVPEDRHRDGLVLDMTMAENLAL 369

LIQAITGLRKV SG I+IK +VT SSR+ITEL SVGHVPEDRHRDGL+LD+++AEN AL

Sbjct: 306 LIQAITGLRKVKSGSIMIKNEVTHLSSRKITEL SVGHVPEDRHRDGLILDLSLAENTAL 365

65

Query: 370 QTYTKEPLSHKGILNFAKIKEYARQLMTEFDVRGAGEHVLARGFSGGNQQAIIAREVDR 429

QTYT+PLS GILN+ KI +YARQLM EFDVGA E V ARGFSGGNQQAIIAREVDR

Sbjct: 366 QTYTQPLSQNGILNVTYKINDYARQLMKEFDVRGANELVPARGFSGGNQQAIIAREVDR 425

Query: 430 DPDLLIVSQPTRGLDVGAI EYIHKRLIEERDKGKAVLVVSFELDEILNLSDRIAVIHDGK 489

-1686-

DPDLLIVSQPTRGLDVGAI EYIHKRLI+ERDKGKAVLVVSFELDEILNLSDR IAVIHDGK
 Sbjct: 426 DPDLLIVSQPTRGLDVGAI EYIHKRLIKERDKGKAVLVVSFELDEILNLSDR IAVIHDGK 485

Query: 490 IQGIVKPDQTNKQELGILMAGGKIEKEERDV 520
 IQGIV P+ TNKQELGILMAGG I KEE V
 Sbjct: 486 IQGIVSPENTNKQELGILMAGGSIHKEEGHV 516

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 1521

A DNA sequence (GBSx1612) was identified in *S.agalactiae* <SEQ ID 4679> which encodes the amino acid sequence <SEQ ID 4680>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15143 GB:Z99120 similar to ABC transporter (lipoprotein)
 [Bacillus subtilis]
 Identities = 164/335 (48%), Positives = 224/335 (65%), Gaps = 10/335 (2%)

Query: 18 LAACGHRGASKSGGKS-DSLKVAMVTD TGGVDDKSFNQSGWEGMQAWGKKNGLKKG A-GF 75
 L ACG+ S G+ + VAMVTD GGVDDKSFNQSG WEG+QA+GK+NGLKKG G+
 Sbjct: 11 LGACGNSEKSSSGSGEGKNKFSVAMVTDVGGVDDKSFNQSAWEGIQAFGKENGKKGKNGY 70

Query: 76 DYFQSASESDYATNLD TAVSSGYKLIFGIGFSLHDAIDKAADNNKDVNYVIVDDVIKGD 135
 DY QS S++DY TNL+ + LI+G+G+ + D+I + AD K+ N+ I+D V+ KD
 Sbjct: 71 DYLSQSKSDADYTTN LNKLARENFDLIYGVGYLMEDSISEIADQRKNTNF AIDAVVD-KD 129

Query: 136 NVASVVFADNESAYLAGIAAAKTTKT KT VGFVGMSESEVITRFEKGFEAGVKSVDKSIKI 195
 NVAS+ F + E ++L G+AAA ++K+ +GFVGMSESE+I +FE GF AGV++V+ +
 Sbjct: 130 NVASITFKEQEGSFLVGVAALSSKSGKIGFVGMSESELIKKEFEVGFAGVQAVNPKAVV 189

Query: 196 KVDYAGSFGDAAGKKTIAAAQYASGADIVYQVAGGTGAGVFEAKSRNESLKEADKVWVL 255
 +V YAG F A GK A + Y SG D++Y AG TG GVF+EAK+ + + D VWV+
 Sbjct: 190 EVKYAGGFDKADVGKATAESMYKSGVDVIYHSAGATGTGVFTEAKNLKKEDPKRD-VWVI 248

Query: 256 GVD RDQA AEGKYTSKDGKASNFVLASSIKEVGSVELIATKTSKGKFPGGNVTTYGLKD G 315
 GVD+DQ AEG+ +G N L S +K+V VE + K S GKFPGG TYGL
 Sbjct: 249 GVDKDQY AEGQV---EGTDDNVTLTSMVKVDTVVEDVTKKASDGKFPGGETLTYGLDQD 305

Query: 316 GV DIATT--NLSDDAVKAIKEAKAKIISGDIKVP S 348
 GV I+ + NLSDD +KA+ + K KII G +++P+
 Sbjct: 306 GVGISPSKQNLSD DVIKAVDKWKKKIIDG-LEIPA 339

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 861> which encodes the amino acid sequence <SEQ ID 862>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-1687-

Identities = 275/351 (78%), Positives = 312/351 (88%), Gaps = 3/351 (0%)

Query: 1 MNKKIAGIGLASIAVLSLAACGHRGASKSG--GKSDSLKVAMVTDGTVDDKSFNQSGWE 58
 MNKK G+GLAS+AVLSLAACG+RGASK G GK+D LKVAMVTDGTVDDKSFNQSG WE
 5 Sbjct: 1 MNKKFIGLGLASVAVLSLAACNKGASKGASGKTD-LKVAMVTDGTVDDKSFNQSAWE 59

Query: 59 GMDQWGGKNGGLKKGAGFDYFQSASESDYATNLDIAVSSGYKLIFGIGFSLHDAIDKAADN 118
 G+Q+WGK+ GL+KG GFDYFQS SES+YATNLDIAVS GY+LI+GIGF+L DAI KAA +
 10 Sbjct: 60 GLQSWGKEMGLQKGTGFDYFQSTSESEYATNLDIAVSSGGYQLIYGIGFALKDAIAKAAGD 119

Query: 119 NKDVNYVIVDDVIKGDNVASVVFADNESAYLAGIAAAKTTKTKTVGFGVGGMESEVITRF 178
 N+ V +VI+DD+I+GKDNVASV FAD+E+AYLAGIAAAKTTKTKTVGFGVGGME VITRF
 15 Sbjct: 120 NEGKVFVIIDDIIEGKDNVASVTFADHEAAYLAGIAAAKTTKTKTVGFGVGGMEGTVITRF 179

Query: 179 EKGFEAGVKSVDKSIKIKVDYAGSFGDAAGKGTIAAAQYASGADIVYQVAGGTGAGVFSE 238
 EKGFEAGVKSVD +I++KVDYAGSFGDAAGKGTIAAAQYA+GAD++YQ AGGTGAGVF+E
 20 Sbjct: 180 EKGFEAGVKSVDVTIQKVDYAGSFGDAAGKGTIAAAQYAAGADVIYQAGGTGAGVFNE 239

Query: 239 AKSRNESLKEADKVVVLGVDRDQAAGKGTISKDGKASNFVLASSIKEVGKSVELIATKTS 298
 AK+ NE EADKVVV+GVDRDQ EGKYSKDGK +NFVLASSIKEVGK+V+LI + +
 25 Sbjct: 240 AKAIKEKRSEADKVVVIGVDRDQKDEGKYSKDGKASNFVLASSIKEVGKAVQLINKQVA 299

Query: 299 KGKFPGGNVTTYGLKDGGVDIATTNLSDDAVKAIKEAKAKIISGDIKVP SK 349
 KFPGG T YGLKDGGV+IATTN+S +AVKAIKEAKAKI SGDIKVP K
 30 Sbjct: 300 DKKFPGGKTTVYGLKDGGVETATTNVSKAVKAIKEAKAKISGDIKVPEK 350

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9061> which encodes amino acid sequence <SEQ ID 9062>. Analysis of this protein sequence reveals the following:

Possible site: 17
 30 >>> May be a lipoprotein

----- Final Results -----
 35 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 414 bits (1052), Expect = e-117
 40 Identities = 196/347 (56%), Positives = 253/347 (72%), Gaps = 2/347 (0%)

Query: 1 MNKKVMSLGLVSTALFTLGGCTNNNAKQT--TDNSLKIAMITNQTGIDDKSFNQSAWEGL 58
 MNKK+ +GL S A+ +L C + A ++ +SLK+AM+T+ G+DDKSFNQSG WEG+
 45 Sbjct: 1 MNKKIAGIGLASIAVLSLAACGHRGASKSGKSDSLKVAMVTDGTVDDKSFNQSGWEGM 60

Query: 59 QAWGKENKLEKKGKGYDYFQSANESEFTTNLESATVNGYNLVFGIGFPLHDAVEKVAANNP 118
 QAWGK+N L+KG G+DYFQSA+ES++ TNL++AV++GY L+FGIGF LHDA++K A NN
 50 Sbjct: 61 QAWGKNGGLKKGAGFDYFQSASESDYATNLDIAVSSGYKLIFGIGFSLHDAIDKAADNNK 120

Query: 119 DNHFAIVDDVIKGDKNVASITFSDEAAAYLAGVXXXXXXXXXXQVGVGGMEGDVVKRFK 178
 D ++ IVDDVIK+ NVA+ F+D+E+AYLAG+ VGFVGGME +V+ RFEK
 55 Sbjct: 121 DVNYVIVDDVIKGDKNVASVVFADNESAYLAGIAAAKTTKTKTVGFGVGGMESEVITRFK 180

Query: 179 GFEAGVKSVDVTIKVRVAYAGSFXXXXXXXXXXXXXXXXXXEGADVYHAAGGTGAGVFSEAK 238
 GFEAGVKSVD +IK++V YAGSF GAD++Y AGGTGAGVFSEAK
 60 Sbjct: 181 GFEAGVKSVDKSIKIKVDYAGSFGDAAGKGTIAAAQYASGADIVYQVAGGTGAGVFSEAK 240

Query: 239 SINEKRKEEDKVVVIGVDRDQSEDKYTTKDGKSNFVLTSSSIKEVGKALVKVAVKTSK 298
 S NE KE DKVVV+GVDRDQ+ +GKYT+KDGK++NFVL SSIKEVGK++ +A KTS+
 65 Sbjct: 241 SRNESLKEADKVVVLGVDRDQAAGKGTISKDGKSNFVLASSIKEVGKSVELIATKTSK 300

Query: 299 QFFGGQITTFGLKEGGVSLTTDALTDQTXXXXXXXXXXXXXXTITVP 345
 +FPGG +TT+GLK+GGV + T L+ D G I VP
 Sbjct: 301 KFPGGNVTTYGLKDGGVDIATTNLSDDAVKAIKEAKAKIISGDIKVP 347

-1688-

SEQ ID 4680 (GBS211) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 6; MW 40kDa).

The GBS211-His fusion product was purified (Figure 205, lane 8) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 259A) and FACS (Figure 259B). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1522

A DNA sequence (GBSx1613) was identified in *S.agalactiae* <SEQ ID 4681> which encodes the amino acid sequence <SEQ ID 4682>. This protein is predicted to be cytidine deaminase (cdd). Analysis of this protein sequence reveals the following:

Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2112(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9801> which encodes amino acid sequence <SEQ ID 9802> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB51906 GB:AJ237978 cytidine deaminase [Bacillus psychrophilus]
Identities = 66/114 (57%), Positives = 81/114 (70%)

Query: 26 KASENAYVPYSKFPVGAALRTAEGKIFTGCNVENISYGLANCAERTAIFKAVSEGYKDFS 85
KA E AYVPYSKFPVGAAL +G I+ GCN+EN +Y + NCAERTA FKA VS+G + F
Sbjct: 12 KAREQAYVPYSKFPVGAALLAEDGTIYHGCNIENSAYSMINCAERTAFFKAVSDGVRSFK 71

Query: 86 EIAIYGNTEEPISPCGACRQVMVEFFNKNKAVTTLIAKNGKTIVETTVGELLPSYF 139
+A+ +TE P+SPCGACRQV+ EF N + V L G ETTV +LLP +F
Sbjct: 72 ALAVVADTEGPVSPCGACRQVIAEFCNGSMPVYLTLNLKGDIEETTVAKLLPGAF 125

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4683> which encodes the amino acid sequence <SEQ ID 4684>. Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0041(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB15143 GB:Z99120 similar to ABC transporter (lipoprotein)
[Bacillus subtilis]
Identities = 152/339 (44%), Positives = 223/339 (64%), Gaps = 11/339 (3%)

Query: 8 LGLVSTALFTLGGCTNN---SAKQTTDNLKIAMITNQTGIDDKSFNQSAWEGQLQAWGKE 64
+ LV A LG C N+ S N +AM+T+ G+DDKSFNQSAWEG+QA+GKE
Sbjct: 1 MSLVIAAGTILGACGNSEKSSGSGEGKNKFSVAMVTDVGGVDDKSFNQSAWEGIQAQFGKE 60

-1689-

Query: 65 NKLEKKG-GYDYFQSANESSEFTTNLESASVINGYNLVFGIGFPLHDAVEKVAANNPDNHFA 123
 N L+KKG GYDY QS +++++TTNL ++L++G+G+ + D++ ++A + +FA
 Sbjct: 61 NGLKKKGNGYDYLOSKSDADYTTNLNKLARENFDLIYGVGYLMEDSISEIADQRKNTNFA 120

5 Query: 124 IVDDVIKQKQNVASITFSDEHAAYLAGVAAAKTTKTQVGFVGGMEGDVVKRFEKGFEG 183
 I+D V+ + NVASITF + E ++L GVAAA ++K+ ++GFVGGME +++K+FE GF AG
 Sbjct: 121 IIDAVVD-KDNVASITFKEQEGSFLVGVAAALSSKSGKIGFVGGMESELIKKEFEVGFAG 179

10 Query: 184 VKSVDDTIKVRVAYAGSFADAAGKGTIAAAQYAGADVIYHAAGGTGAGVFSEAKSINEK 243
 V++V+ V V YAG F A GK A + Y G DVIYH+AG TG GVF+EAK++ ++
 Sbjct: 180 VQAVNPKAVVEVKYAGGFDKADVGKATAESMYKSGVDVIYHSAGATGTGVFTEAKNLKKE 239

15 Query: 244 RKEEDKVVWVIGVDRDQSEDGKYTTTGKGSANFVLTSSIKEVGKALVKVAVKTSQDFPGG 303
 + D VVWIGVD+DQ +G+ +G N LTS +K+V + V K S+ +FPGG
 Sbjct: 240 DPKRD-VWVIGVDKQYAEQGV---EGTDDNVTLTSMVKKVDTTVVEDVTKKASDGKFPFGG 295

20 Query: 304 QITTFGLKEGGVSLTTDA--LTQDTKKAIEAAKKAIEG 340
 + T+GL + GV ++ L+ D KA++ KK II+G
 Sbjct: 296 ETLTFGLDQDGVGISPSKQNLSDVIKAVDKWKKIIDG 334

An alignment of the GAS and GBS proteins is shown below.

Identities = 88/128 (68%), Positives = 107/128 (82%)

25 Query: 15 MGNIELKKLAVKASENAYVPYSKFPVGAALRTAEGKIFTGCNVENISYGLANCAERTAIF 74
 MG +L AV+ASE AYVPYS FPVGAAL+T +G I+TGCN+EN+S+GL NC ERTAIF
 Sbjct: 1 MGTDDLVSQAVQASEYAYVPYSHFPVGAALKTGDTIYTGCTNIENVSFGLTNCGERTAIF 60

30 Query: 75 KAVSEGYKDFSEIAIYGNTEEPISPCGACRQVMVEFFNKNKAVTLIAKNGKTVETTVGEL 134
 KA+S+G+K+ EIAIYG T +P+SPCGACRQVM EFF+ ++ VTLLIAKNG+TVE TVG+L
 Sbjct: 61 KAISDGHKELVEIAIYGETMQFVSPCGACRQVMAEFFDPSSLVTLLIAKNGQTVEMTVGDL 120

35 Query: 135 LPYSFVDL 142
 L YSF DL
 Sbjct: 121 LLYSFIDL 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1523

40 A DNA sequence (GBSx1614) was identified in *S.agalactiae* <SEQ ID 4685> which encodes the amino acid sequence <SEQ ID 4686>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2979(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 A related GBS nucleic acid sequence <SEQ ID 9799> which encodes amino acid sequence <SEQ ID 9800> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:CAB11882 GB:Z99104 alternate gene name: ybaA~similar to
 hypothetical proteins [Bacillus subtilis]
 Identities = 90/201 (44%), Positives = 144/201 (70%), Gaps = 5/201 (2%)
 Query: 1 MANMYTENPNVEHDIHELNVKLLGESFSFLTDAGVFSKRMIDYGSQVLLNSLHF-EKNK 59
 M+ YY+E P+V+ + + +L + F+F +D+GVFSK+ +D+GS++L++S E
 Sbjct: 1 MSEHYYSKPSVKSXKQTSFRLRNKDFTFTSDSGVFSKKEVDGSRLLIDSFEPEVEG 60

-1690-

Query: 60 SLLDLGCGYGPLGISLAK-VQGVKATMVDINTRALELAKKNATRNGVV-VEVFQSNIIYEN 117
 +LD+GCGYGP+G+SLA + M+D+N RA+EL+ +NA +NG+ V+++QS+++ N
 Sbjct: 61 GILDVCGYGFPLGISLASDFKDRITIHMIDVNERAVELSNENAEQNGITNVKIYQSDLFEN 120

5 Query: 118 I--SKTFDYIISNPPIRAGKQVVHSIIIESICYLNTGGSLTIVIQKQKQAPSASAKMLDT 175
 + ++TF I++NPPIRAGK+VVH+I E+S +L G L IVIQKQKQAPSA K+ +
 Sbjct: 121 VDSAQTFASILTNPPIRAGKKVVHAI FEKSAEHLKASGELWTVIQKQKQAPSASIEKLEEL 180

10 Query: 176 FGNC DILKKDKGYIILRSEKV 196
 F +++K KGYI++++KV
 Sbjct: 181 FDEVSVVQKKKGYIILKAKKV 201

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4687> which encodes the amino acid sequence <SEQ ID 4688>. Analysis of this protein sequence reveals the following:

15 Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.4232(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 139/195 (71%), Positives = 165/195 (84%)

25 Query: 1 MANMYYTENPNVEHDIHNLVKGESFSFLTDAGVFSKRMIDYGSQVLLNSLHFEKNKS 60
 M MYE ENP+ HDIHE+ V+LL F+FLTD+GVFSK+M+D+GSQVLL +L+F++N+
 Sbjct: 12 MTKMYDENPDSLHDIHEVKVELLNHPFTFLTDGVSFVKRMVDFGSQVLLKTLNFKENER 71

30 Query: 61 LLDLGCGYGPLGISLAKVQGVKATMVDINTRALELAKKNATRNGVVVEVFQSNIIYENISK 120
 +LDLGCGYGPLGISLAKVQ V AT+VDIN RAL+LA+KNAT N V V +FQSNIIYENIS
 Sbjct: 72 VLDLGCGYGPLGISLAKVQRVDATLVDINNRLDLARKNATNNQVAVTIFQSNIIYENISG 131

35 Query: 121 TFDYIISNPPIRAGKQVVHSIIIESICYLNTGGSLTIVIQKQKQAPSASAKMLDTFGNCD 180
 F++IISNPPIRAGK+VVHSIIE+SI +L G L TIVIQKQKQAPSASAKM FGN +
 Sbjct: 132 HFEHIISNPPIRAGKRVVHSIIEKSIDFLVNGDLTIVIQKQKQAPSASAKMATIFGNVE 191

40 Query: 181 ILKKDKGYIILRSEK 195
 IL+KDKGYI+LRS K
 Sbjct: 192 ILRKDKGYIVLRSIK 206

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1524

45 A DNA sequence (GBSx1615) was identified in *S.agalactiae* <SEQ ID 4689> which encodes the amino acid sequence <SEQ ID 4690>. This protein is predicted to be pantothenate kinase (coaA). Analysis of this protein sequence reveals the following:

50 Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5021(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06594 GB:AP001516 pantothenate kinase [Bacillus halodurans]
 Identities = 140/307 (45%), Positives = 203/307 (65%), Gaps = 5/307 (1%)

-1691-

Query: 4 EFINFDRISRENWKDLHQSQALLTEKELESIKSLNDNINIQDVIDIYLPLINLIQIYKR 63
 +F + +SR WK L + S + E+ELE + LN+ I + +V DIY+PL L+ ++
 Sbjet: 8 DFFPYTVLSRSQWKSRLKASSLPINEQEQLVGLNEPITLNEVADITVPLAELLHVHAT 67

Query: 64 SQENLSFSKAIFLKKENYQRPFIIGISGSVAVGKSTTSRLQLLISRTFKDSHVELVTTD 123
 + + L K F + PFIIG++GSVAVGKSTT+RLLQ L+ + HV+LVTTD
 Sbjet: 68 AYQRLQQQKRGFFHHGKNRSPFIIGLAGSVAVGKSTTARLLQKLLKAWPEHHHVDLVTTD 127

Query: 124 GFLYPNEKLIQNGILNRKGFPEPYDMESELLNFDITKNGIT-AKIPIYSHEIYDIVPNQL 182
 GFLYPNE L G++++KGFPEPYD+ +L+ FL +K G K P+YSH Y+IV
 Sbjet: 128 GFLYPNETLEARGLMDKKGFPEPYDLEALIRFLSDVKAGEPYVKAPVYSHLTYNIVEGDY 187

Query: 183 QTITPDLFLEGINVFQ-NQQNHL---YMDYFDFSIIYIDAENKQIEEWYLRFNLSLL 238
 Q + PD +I+EGINV Q N++NH + +++D+DFSIIY+DA+ +QI +WY++RF L
 Sbjet: 188 QVVHEPDIVIVEGINVLQVNRNHHIPNVFVSDFDFSIIYVDAKEEQILQWYIERFKLLQ 247

Query: 239 QLAADPSNYHKTQIPPHKAMELAKDIWKTINLVNLEKYIEPTRNRADFIHKGKHHK 298
 A DP++Y+H+F + +A + A IWK IN VNL + I PT++RAD ++ KG HH
 Sbjet: 248 NTAQDPNSYFHRFRHLSEVAEQFATSIWKINGVNLHENILPTKHRADLVLOKGPHEF 307

Query: 299 IDEIYLK 305
 IDE+ L+
 Sbjet: 308 IDEVKLR 314

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4691> which encodes the amino acid sequence <SEQ ID 4692>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4790 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 219/306 (71%), Positives = 269/306 (87%)

Query: 1 MNNEFINFDRISRENWKDLHQSQALLTEKELESIKSLNDNINIQDVIDIYLPLINLIQI 60
 M+NEFINF++ISRE+WK LHQ+++ALLT++EL+SI SLNDNI+I DVIDIYLPLINLIQ+
 Sbjet: 1 MSNEFINFEKISRESWKTLHQKAKALLTQEELKSITSLNDNISINDVIDIYLPLINLIQV 60

Query: 61 YKRSQENLSFSKAIFLKKENYQRPFIIGISGSVAVGKSTTSRLQLLISRTFKDSHVELV 120
 YK +QENLSFSK++FLKK+ RPFIIIGISGSVAVGKSTTSRLQLL+SRT +S VELV
 Sbjet: 61 YKLAQENLSFSKSLFLKKDIQLRPFIIIGISGSVAVGKSTTSRLQLLRSRTHPNSQVELV 120

Query: 121 TTDGFLYPNEKLIQNGILNRKGFPEPYDMESELLNFDITKNGITAKIPIYSHEIYDIVPN 180
 TTDGFLYPN+ LI+ G+LNRKGFPEPY+ME LL+FLD+IKNG TA P+YSH+IYDI+PN
 Sbjet: 121 TTDGFLYPNQFLIEQGGLNLRKGFPEPYDMESELLNFDITKNGITAFAPVYSHDIYDIIPN 180

Query: 181 QLQTIETPDLFLEGINVFQNNQNRHLYMDYFDFSIIYIDAENKQIEEWYLRFNLSLLQL 240
 Q Q+ PDLFI+EGINVFQNNQNRHLYM+DYFDFSIIYIDA++ IE WY++RF S+L+L
 Sbjet: 181 QKQSFNNPDFLIVEGINVFQNNQNRHLYMSDYFDFSIIYIDADSSHIETWYIERFLSILKL 240

Query: 241 AEADPSNYHKTQIPPHKAMELAKDIWKTINLVNLEKYIEPTRNRADFIHKGKHHKID 300
 A+ DP NYY ++ Q+P +A+ A+++WKT+NL NLEK+IEPTRNRA+ I+HK HKID
 Sbjet: 241 AKRDPHNYAQAQLPRSEAIAPARNVWKTVNLENLEKFIETPRNRAELILHKSADHKID 300

Query: 301 EIYLKK 306
 EIYLKK
 Sbjet: 301 EIYLKK 306

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1692-

Example 1525

A DNA sequence (GBSx1616) was identified in *S.agalactiae* <SEQ ID 4693> which encodes the amino acid sequence <SEQ ID 4694>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 59
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.3866(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:BAB05058 GB:AP001511 ribosomal protein S20 (BS20) [Bacillus halodurans]
   Identities = 47/86 (54%), Positives = 59/86 (67%), Gaps = 7/86 (8%)

   Query: 3  VKTLANIKSAIKRAELNVKQNEKNSAQKSAMRTAIKAFEANPSEELYRA----ASSS 55
             +K ANIKSAIKR + N K+ +N++ KSA+RTAIK FEA N E +A A+
   Sbjct: 1  MKGNANIKSAIKRVKTNEKRRIQNASVKSALRTAIKQFEAKVENNDAAEAKAAFVEATKK 60

20  Query: 56  IDKAAASKGLIHTNKASRDKARLATKL 81
             +DKAA+KGLIH N ASR K+RLA KL
   Sbjct: 61  LDKAANKGLIHKNAASRQKSRLAKKL 86

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4695> which encodes the amino acid sequence <SEQ ID 4696>. Analysis of this protein sequence reveals the following:

```

25  Possible site: 59
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
30  bacterial cytoplasm --- Certainty=0.3872(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

35  Identities = 76/82 (92%), Positives = 78/82 (94%)

   Query: 1  MEVKTLANIKSAIKRAELNVKQNEKNSAQKSAMRTAIKAFEANPSEELYRAASSSIDKAA 60
             +EVKTLANIKSAIKRAELNVK NEKNSAQKSAMRTAIKAFEANPSEEL+RAASSSIDKA
40  Sbjct: 1  LEVKTLANIKSAIKRAELNVKANEKNSAQKSAMRTAIKAFEANPSEELFRAASSSIDKAE 60

   Query: 61  SKGLIHTNKASRDKARLATKLG 82
             SKGLIH NKASRDKARLA KLG
   Sbjct: 61  SKGLIHKNKASRDKARLAAKLG 82

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1526

A DNA sequence (GBSx1617) was identified in *S.agalactiae* <SEQ ID 4697> which encodes the amino acid sequence <SEQ ID 4698>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 48
   >>> Seems to have no N-terminal signal sequence
       INTEGRAL    Likelihood =-10.99    Transmembrane    31 - 47 ( 25 - 51)

   ----- Final Results -----
55  bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-1693-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAC35851 GB:AF086736 amino acid-binding protein Abp
[Streptococcus uberis]
Identities = 169/269 (62%), Positives = 203/269 (74%), Gaps = 2/269 (0%)

Query: 29 KNILLTIFGLFMILSACGMSNEMAGIDNWEHYQKEKITIGFDNTFVPMGFESRSGD 88
K ILLT + + L ACG S+ A D W+ Y+KEK IT+GFDNTFVPMGF+ SG

10 Sbjct: 4 KKILLTTLALASTLFLVACGKSSA--AKTDQWDITYKKEKSITLGFNTFVPMGFKDES GK 61

Query: 89 YTGFDIDLANAVFKEYGISVKWQPINWDMKETELNNGNIDLIWNGYSKTAERAKKVAFIN 148
TGFD++LA AVF+EYGI VK+QPINWD+KETEL NG ID+IWNIGYS T ER KVAFA+

15 Sbjct: 62 NTGFDVELAKAVFQYEGIKVKFQPINWDLKETELKNGKIDMIWNGYSVTKERQAKVAFST 121

Query: 149 PYMNNHQVIVTKTSSHSINIKDMKGKGLGAQSGSSGFDAPNAKPDILKKFVKGEAVQYD 208
PYM N QV+VTK SS+I S MKGK LGAQSGSSG+DAF + P +LK VK +A QY+

Sbjct: 122 PYMKNEQVLVTKSSNITSFAAMKGKVLGAQSGSSGYDAFTSNPKVLKDIVKDNDAQYE 181

20 Query: 209 TFTQALIDLKNNRIDGLLIDVYANYYLKQEGNIKAYYFVKTAQGENFVVGARKVDRL 268
TF QA IDLKN+RIDGLLID+VYANYYLKQEG + Y VK+ + GE+F VG RK D+ L

Sbjct: 182 TFIQAFIDLKNDRIDGLLIDKVVYANYYLKQEGELTNYNIVKSEFDGEDFAVGVRKEDKIL 241

Query: 269 IEKINKAFKQLHNKGRFQKISYKWFGEDEV 297
++ IN AF +L+ G+FQ+IS KWFEDEV

25 Sbjct: 242 LKNINSAFTKLYKTGKFQEI SQKWFGEDEV 270

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4699> which encodes the amino acid sequence <SEQ ID 4700>. Analysis of this protein sequence reveals the following:

30 Possible site: 21

>>> May be a lipoprotein

----- Final Results -----

35 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

40 >GP:AAC35851 GB:AF086736 amino acid-binding protein Abp
[Streptococcus uberis]
Identities = 176/277 (63%), Positives = 220/277 (78%), Gaps = 1/277 (0%)

Query: 1 MIICKRTVAILAIASSFFLVACQATKSLKSGDAWGVYQKQKSITVGFNTFVPMGYKDES 60
M +KK + LA+AS+ FLVAC + + K+ D W Y+K+KSIT+GFDNTFVPMG+KDES

45 Sbjct: 1 MNLKKILLTTLALASTLFLVACGKSSAAKT-DQWDITYKKEKSITLGFNTFVPMGFKDES 59

Query: 61 GRCKGFDIDLAKVHFQYGLKVNFAINWDMKEAELNNGKIDVIWNGYSITKERQDKVAF 120
G+ GFD++LAK VF +YG+KV FQ INWD+KE EL NGKID+IWNIGYS+TKERQ KVAFA

50 Sbjct: 60 GKNTGFDVELAKAVFQYEGIKVKFQPINWDLKETELKNGKIDMIWNGYSVTKERQAKVAF 119

Query: 121 TDSYMRNEQIIVVKRSDIKTISDMKHKVLGAQSASSGYDSLRLTPKLLKDFIKNKDANQ 180
+ YM+NEQ++V KK S+I + + MK KVLGAQS SSGYD+ PK+LKD +K+ DA Q

55 Sbjct: 120 STPYMKNEQVLVTKSSNITSFAAMKGKVLGAQSGSSGYDAFTSNPKVLKDIVKDNDAQ 179

Query: 181 YETFTQAFIDLKSDRIDGILIDKVVYANYYLAKEGQLENYRMIPTTFENEAFSVGLRKEDK 240
YETFT QAFIDLK+DRIDG+LIDKVVYANYYL +EG+L NY ++ + F+ E F+VG+RKEDK

Sbjct: 180 YETFTQAFIDLKNDRIDGLLIDKVVYANYYLKQEGELTNYNIVKSEFDGEDFAVGVRKEDK 239

60 Query: 241 TLQAKINRAFRVLYQNGKFQAISEKWFGEDEVATANIK 277
L IN AF LY+ GKFK IS+KWFGEDEVATENVK

Sbjct: 240 ILLKNINSAFTKLYKTGKFQEI SQKWFGEDEVATENVK 276

-1694-

An alignment of the GAS and GBS proteins is shown below.

Identities = 151/266 (56%), Positives = 189/266 (70%), Gaps = 4/266 (1%)

```

5  Query: 32  LLTIIFGLFMILSACGMSNKEMAGIDNWEHYQKEKKITIGFDNTFVPMGFESRSGDYTG 91
   +L I   F++  AC + K +   D W  YQK+K IT+GFDNTFVPMG++ SG  G
Sbjct: 10  ILAIASSFFLV---AC-QATKSLKSGDAWGVYQKQKSITVGFDNTFVPMGYKDESGRCKG 65

10 Query: 92  FDIDLANAVFKEYGISVKWQPINWDMKETELNNGNIDLIWNGYSKTAERAKKVAFTNPYM 151
   FDIDLA VF +YG+ V +Q INWDMKE ELNNG ID+IWNNGYS T ER  KVAFT+ YM
Sbjct: 66  FDIDLAKSVFHQYGLKVNFOAINWDMKEAELNNGKIDVIWNGYSITKERQDKVAFTDSYM 125

15 Query: 152 NNHQVIVTKTSSHINSIKDMKGGKLGQAQSGSSGFDAPNAKPDILKKFVKGKEAVQYDTFT 211
   N Q+IV K S I +I DMK K LGAQS SSG+D+   P +LK F+K K+A QY+TFT
Sbjct: 126 RNEQIIVVKRSDIKTISDMKHKVLGAQSASSGYDSLRLTPKLLKDFIKNKDANQYETFT 185

20 Query: 212 QALIDLKNNRIDGLLIDEVYANYYLKQEGNIKAYYFVKTAQGENFVVGARKVDRRLIEK 271
   QA IDLK++RIDG+LID+VYANYYL +EG ++ Y + T ++ E F VG RK D+ L K
Sbjct: 186 QAFIDLKSDRIDGILIDKVYANYYLAKEGQLENYRMIPTTFENEAFSVGLRKEDKTLQAK 245

25 Query: 272 INKAFKQLHNKGRFQKISYKWFGEDEV 297
   IN+AF+ L+  G+FQ IS KWFG+DV
Sbjct: 246 INRAFRVLYQNGKFQAISEKWFGDDV 271

```

25 A related GBS gene <SEQ ID 8833> and protein <SEQ ID 8834> were also identified. Analysis of this protein sequence reveals the following:

```

30 Lipop Possible site: 22  Crend: 4
   Sequence Pattern: CGMS
   SRCFLG: 0
   McG: Length of UR: 22
   Peak Value of UR: 3.05
   Net Charge of CR: 2
   McG: Discrim Score: 11.16
   GvH: Signal Score (-7.5): -1.96
   Possible site: 24
35 >>> May be a lipoprotein
   Amino Acid Composition: calculated from 23
   ALOM program count: 0 value: 8.96 threshold: 0.0
   PERIPHERAL Likelihood = 8.96 68
   modified ALOM score: -2.29
40
   *** Reasoning Step: 3
   ----- Final Results -----
45 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

50 62.2/75.8% over 270aa
   Streptococcus uberis
   GP|3603430| amino acid-binding protein Abp Insert characterized

   ORF00904(385 - 1203 of 1503)
55 GP|3603430|gb|AAC35851.1||AF086736(4 - 274 of 277) amino acid-binding protein Abp
   {Streptococcus uberis}
   %Match = 34.8
   %Identity = 62.1 %Similarity = 75.7
   Matches = 169 Mismatches = 65 Conservative Sub.s = 37

60 153      183      213      243      273      303      333      363
   FHYLGGSNVSH*LVR**LIHRLLVMSQALALIQCCKK*KN*FYKIEKQV*HKL**HMI FNLLKVYLIRFSKLLLSRL

   393      423      453      483      513      543      573      603
   GGRLLTHKNILLTIIFGLFMILSACGMSNKEMAGIDNWEHYQKEKKITIGFDNTFVPMGFESRSGDYTGFDIDLANAVF

```


-1695-

```

      : | | | | :      : | | | | : | | | : | | | | | | | | : | | | | | : | | |
      MNLKKILLTTLALASTLFLVACGKSS--AAKTDQWDYKKEKSITLGFDFVPMGFKDESGKNTGFDVELAKAVF
      10      20      30      40      50      60      70

5      633      663      693      723      753      783      813      843
      KEYGISVKWQPINWDMKETELNNGNIDLIWNGYSKTAERAKKVAFTNPYMNHQQVIVTKTSSHINSIKDMKGKILGAQSG
      : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
      QEYGIKVKFQPINWDLKETELKNGKIDMIWNGYSVTKERQAKVAFSTPYMKNEQVLVTKSSNITSFAAMKGVILGAQSG
      90      100      110      120      130      140      150

10     873      903      933      963      993      1023      1053      1083
      SSGFDFNAKPDILKKFVKGEAVQYDTFTQALIDLKNNRIDGLLIDEVYANYYLKQEGNIKAYYFVKTAQQENFVVG
      | | : | | : | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
      SSGYDAFTSNPKVLKDIVKNDATQYETFIQAFIDLKNDRIDGLLIDKVYANYYLKQEGELTNYNIVKSEFDGEDFAVGV
      170      180      190      200      210      220      230

15     1113      1143      1173      1203      1233      1263      1293      1323
      RKVDRRLIEKINKAFKQLHNKGRFQKISYKWFGEDEVYSKE*KTRNFS*SFILRKN*IKNIDISDVF*VN*PSLVARRALS
      | | : | | : | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
      RKEDKILLKNINSAFTKLYKTGKFQETSQKWFGEDEVATENVKK
      250      260      270

```

SEQ ID 8834 (GBS225) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 10; MW 32kDa). The GBS225-His fusion product was purified (Figure 205, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 266), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1527

A DNA sequence (GBSx1618) was identified in *S.galactiae* <SEQ ID 4701> which encodes the amino acid sequence <SEQ ID 4702>. This protein is predicted to be arginine ABC transporter, ATP-binding protein (glnQ). Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3229(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAB49429 GB:U73111 glutamine transport ATP-binding protein GLNQ
      [Salmonella typhimurium]
      Identities = 94/210 (44%), Positives = 146/210 (68%), Gaps = 3/210 (1%)

45      Query: 1      MLELNISKCYGQKEIFKDFNLTVEEGKILSLVGPSSGGGKTTLRLMLAGLEKIDSGTIVH 60
      M+E KN+SK +G ++ + +L + +G+++ ++GPSG GK+TLRL + LE+I SG ++
      Sbjct: 1      MIEFKNVSKHFGPTQVLHNIDLNIRQGEVVIIIGPSGSGKSTLLRCINKLEEITSGLDIV 60

50      Query: 61      DGKEVS---VDHLETLNLLGFVFQDFQLFPHLTVLDNLILSPVKTMGLSKELAKEKALVL 117
      DG +V+ VD G VFQ F LPHLT L+N++ P++ G+ KE A+++A L
      Sbjct: 61      DGLKVNDPKVDERLIRQEAGMVFFQFYLFPHLTALENVMFGLPLRVGVKKEAEKQAKAL 120

55      Query: 118     LERLGLKDHALVYFFSLSGGQKQORVALARAMMIDPQIIGYDEPTSALDPELRQVEKILIL 177
      L ++GL + A YP LSGGQ+QRVA+ARA+ + P+++ +DEPTSALDPELR EV K++
      Sbjct: 121     LAKVGLAERAHHPSELSCGQQQRVAIARALAVKPKMMLFDEPTSALDPELRHEVLKVMQ 180

      Query: 178     QNRETGMTQIVVTHDLQFAESISDTILKIN 207

```

-1696-

E GMT ++VTH++ FAE ++ ++ I+
 Sbjct: 181 DLAEEGMTMVIVTHEIGFAEKVASRLIFID 210

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4703> which encodes the amino acid
 5 sequence <SEQ ID 4704>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2146(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 164/209 (78%), Positives = 183/209 (87%)

Query: 1 MLELKNISKCYGQKEIFKDFNLTVEEGKILSLVGPSSGGGKTTLLRMLAGLEKIDSGTIVH 60
 MLELKNISK +GQK IF FNLTV++G++LSLVGPS GGKTTLLRMLAGLE IDSG + +
 20 Sbjct: 1 MLELKNISKQFGQKTIFDGFNLTVDGEVLSLVGPSSGGGKTTLLRMLAGLESIDSGQVVF 60

Query: 61 DGKEVSVVDHLETNLNLLGFVFQDFQLFPHLTVDNLILSPVKTMGLSKELAKEKALVLLER 120
 +G++V +DHLE NLLGFVFQDFQLFPHLTVDNL LSP TMG K AKEKAL LL R
 25 Sbjct: 61 NGEDVGDIDHLENRNLLGFVFQDFQLFPHLTVDNLTLSTITMGKQKADAKEKALDLLAR 120

Query: 121 LGLKDHALVYPFSLSGGQKQVALARAMMIDPQIIGYDEPTSAIDPELRQEVEKLILQNR 180
 LGLK+HA VYP+SLSGGQKQVALARAMMIDPQIIGYDEPTSAIDPELRQ VE LI+QNR
 30 Sbjct: 121 LGLKEHAQVYPYSLSGGQKQVALARAMMIDPQIIGYDEPTSAIDPELRQTVEALIVQNR 180

Query: 181 ETGMTQIVVTHDLQFAESISDTILKINPK 209
 E G+TQIVVTHDL FAE+ISD I+++NPK
 35 Sbjct: 181 EMGITQIVVTHDLVFAEAIISDRIIRVNPK 209

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

35 Example 1528

A DNA sequence (GBSx1619) was identified in *S.agalactiae* <SEQ ID 4705> which encodes the amino
 acid sequence <SEQ ID 4706>. This protein is predicted to be amino acid ABC transporter, permease
 protein (glnP). Analysis of this protein sequence reveals the following:

40 Possible site: 16
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -8.12 Transmembrane 102 - 118 (96 - 120)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9341> which encodes amino acid sequence <SEQ ID 9342>
 was also identified.

50 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA98402 GB:AP002545 ABC amino acid transporter permease
 [Chlamydomonas reinhardtii J138]
 Identities = 55/127 (43%), Positives = 83/127 (65%), Gaps = 1/127 (0%)

55 Query: 3 AAIIAFTMNYAAYFAEIFRGGIESIPKGYEAAKVLKFSKFQTVWYIVLPQVFKIVLPSV 62
 A IIA +MN AAY AE RGGI S+ GQ+E+A VL + K+Q YI+ PQVFK +LPS+

-1697-

Sbjct: 89 AGIIALSMNSAAYLAENIRGGINSLSIGQWESAMVLGYKKYQIFVYIIYPQVFNILPSL 148

Query: 63 FNETITLVKDSLVYILGVGDLLLESKTAANRDATLAPMF-IAGGIYLLIGLLTILSKQ 121

NE ++L+K+SS++ ++GV +L +K +R+ M+ I G+Y L+ + +S+

Sbjct: 149 TNEFVSLIKESSILMVGVPELTQVTKDIVSRELNPMEYLICAGLYFLMTSSFSICISRL 208

Query: 122 VEKRFNY 128

EKR +Y

Sbjct: 209 SEKRRSY 215

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4707> which encodes the amino acid sequence <SEQ ID 4708>. Analysis of this protein sequence reveals the following:

Possible_site: 34

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -11.57 Transmembrane 21 - 37 (7 - 44)

INTEGRAL Likelihood = -10.93 Transmembrane 185 - 201 (178 - 206)

INTEGRAL Likelihood = -3.29 Transmembrane 63 - 79 (62 - 81)

----- Final Results -----

bacterial membrane --- Certainty=0.5628(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB05181 GB:AP001512 ABC transporter (permease) [Bacillus halodurans]
Identities = 86/206 (41%), Positives = 126/206 (60%), Gaps = 1/206 (0%)

Query: 4 IQQVLPSLLDGALVTLQVFFIVIIILSIPLGAILAFLMKIPFKPLQWFLTYVWMMRGTP 63

IQ +P +L+G VTLQ + ++ + LG +LA ++ +WF Y + RGTP

Sbjct: 8 IQPFMPFMLEGVWVTLQFVSLLFLGLVIGIVLAIFKISKYRLFRWFADFYTSIFRGTP 67

Query: 64 ILQLIFFYVLPVSGISFDRMPAAILAFTLNAAAYFAEIPRGGIEAIPKQYEAQVVKL 123

+LQL+ Y LP G+ + AA LAF LN AAY +EI R GI+A+ KQQ EAA+ L +

Sbjct: 68 ILQLMIYLALEQFGVDISQFQAFLAFLGNSAAYVSEIIRAGIQAVDKQREAAEALGI 127

Query: 124 KPLQTIRYIILPQVFKIVLPVFNVINLVKDSLVYVVGVDLL-LASKTAANRDATLA 182

+ IILPQ + +LP++FNE INL K+S++V V+GV DL+ A T+A L

Sbjct: 128 PYRPMMLRIILPQAMRNILPALFNEFINLTKEAIVSVIGVTDLMRRAQITSATYLYLE 187

Query: 183 PMFIAGLIYLLIGLVTIISKQVEKR 208

P+ GLIY +L+ +T+I + +E+R

Sbjct: 188 PLLFVGLIYVVLVGMCLTVIGRLLERR 213

An alignment of the GAS and GBS proteins is shown below.

Identities = 112/130 (86%), Positives = 121/130 (92%)

Query: 1 MPAAIIAFTMNYAAYFAEIPRGGIESIPKQYEAQVVKLFSKFQTVWYIVLPQVFKIVLP 60

MPAAI+AFT+NYAAYFAEIPRGGIE+IPKQYEAQVVKL QT+ YI+LPQVFKIVLP

Sbjct: 84 MPAAI LAFTLNAAAYFAEIPRGGIEAIPKQYEAQVVKLKLKPLQTIRYIILPQVFKIVLP 143

Query: 61 SVFNETITLVKDSLVYILGVGDLLLESKTAANRDATLAPMFIAGGIYLLIGLLTILSK 120

SVFNE I LVKDSLVY+LGVDLLL SKTAANRDATLAPMFIAG IYLLIGL+TI+SK

Sbjct: 144 SVFNEVINLVKDSLVYVVGVDLLLASKTAANRDATLAPMFIAGLIYLLIGLVTIISK 203

Query: 121 QVEKRFNYYK 130

QVEKRFNYY+

Sbjct: 204 QVEKRFNYYQ 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1698-

Example 1529

A DNA sequence (GBSx1620) was identified in *S.agalactiae* <SEQ ID 4709> which encodes the amino acid sequence <SEQ ID 4710>. This protein is predicted to be minidiscs. Analysis of this protein sequence reveals the following:

```

5   Possible site: 61
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -9.66    Transmembrane  44 - 60 ( 39 - 66)
      INTEGRAL    Likelihood = -7.96    Transmembrane 129 - 145 ( 123 - 147)
10   INTEGRAL    Likelihood = -5.15    Transmembrane  13 - 29 (  9 - 33)
      INTEGRAL    Likelihood = -2.39    Transmembrane  94 - 110 ( 94 - 110)

      ----- Final Results -----
                bacterial membrane --- Certainty=0.4864 (Affirmative) < succ>
                bacterial outside  --- Certainty=0.0000 (Not Clear) < succ>
15                  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF49688 GB:AE003532 mnd gene product [Drosophila melanogaster]
Identities = 48/145 (33%), Positives = 78/145 (53%), Gaps = 8/145 (5%)

20   Query: 7   IKQTYGLMTTIAMIVGVVIGSGIYFKVDDILKFTGGDVFLGMVILVLGFSFIVFGSLISIS 66
      +K+  GL+  +A+IVGV++GSGI+      +LKF+ G +   +++ VL      + G+L +
      Sbjct: 39 LKKQIGLLDGVAIIVGVIVGSGIFVSPKGVLFKS-GSIGQSLIVWVLSGVLSMVGALCYA 97

25   Query: 67 ELAIRTSESGGIFSYYEKYVSPALAATLGLFASFLYL-PTLTAIVSWVAIFYTLGE----- 121
      EL      +SGG ++Y      P L A L L+ + L L PT AI +   A Y L
      Sbjct: 98 ELGTMIPKSGGDYAYIGTAFGP-LPAFLYLWVALLILVPTGNAITALTFAIYLLKPFWPS 156

30   Query: 122 -SSSLESQIILAAVYIILSLMNIF 145
      + +E+  +LAA I  L+L+N +
      Sbjct: 157 CDAPIEAVQLLAAAMICVLTILINCY 181

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1530

A DNA sequence (GBSx1621) was identified in *S.agalactiae* <SEQ ID 4711> which encodes the amino acid sequence <SEQ ID 4712>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 21
   >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
                bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
                bacterial outside  --- Certainty=0.0000 (Not Clear) < succ>
45                  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1699-

Example 1531

A DNA sequence (GBSx1622) was identified in *S.agalactiae* <SEQ ID 4713> which encodes the amino acid sequence <SEQ ID 4714>. This protein is predicted to be TRK potassium uptake system protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 27
   >>> Seems to have no N-terminal signal sequence
       INTEGRAL    Likelihood = -0.06    Transmembrane  232 - 248 ( 232 - 248)

10  ----- Final Results -----
       bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
       bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8835> which encodes amino acid sequence <SEQ ID 8836> was also identified. Analysis of this protein sequence reveals the following:

```

15  Lipop: Possible site: -1    Crend: 5
     McG: Discrim Score:      -4.65
     GvH: Signal Score (-7.5): -3.64
     Possible site: 27
20  >>> Seems to have no N-terminal signal sequence
     ALOM program  count: 1 value: -0.06 threshold: 0.0
       INTEGRAL    Likelihood = -0.06    Transmembrane  228 - 244 ( 228 - 244)
       PERIPHERAL  Likelihood = 1.27      428
25  modified ALOM score: 0.51

     *** Reasoning Step: 3

     ----- Final Results -----
30  bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
     bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

35  >GP:AAB90401 GB:AE001046 TRK potassium uptake system protein
     (trkA-2) [Archaeoglobus fulgidus]
     Identities = 136/446 (30%), Positives = 238/446 (52%), Gaps = 12/446 (2%)

Query: 5   MRIIVGGGKVGKTALCRSLVAEKHDVVLIEKKENVLKRVTKQHDIMGIVGNGANYKILEQ 64
Sbjct: 1   MRIVIAGAGEVGHYLAMSLL-APNHDVIIIEKDVSFRFERVSEL-DVVAINGNAANMKVLRD 58

Query: 65  AEVKNCDFIAITDRDEVNMISAVLAKKMGAKETVVRMRNPEYSNPYFKDKNFLGFSSVV 124
Sbjct: 59  A V+ D+F+A+T DEVN++S + AKK+GAK +VR+ NPEY + ++ LG+ ++

45  Query: 125 NPELLAAQYIANTIEFPNATSVVEHFANGRVMLMEFKILEGNKLCHTSMSQIRKKFGNIVI 184
     P+L AQ A I P A V F+ G+V ++E +++EG+K +++ + N+VI
Sbjct: 119 CPQLSLAQEAARLIGIPGAIEVVTTFSGGKVMIEIQVMEGSKADGKAIADLYLP-QNVVI 177

50  Query: 185 CAIERDGLIIPDGDATIQVKDKIFVTGNRIEMILFHNIVKNKVVKNLMVIGAGRIAYYL 244
     +I R+G + IP GD ++ D++ + ++ + V + + + GAG I Y
Sbjct: 178 ASIYRNGHIEIPRGDTVLRAGDRVAIVSKTEDVEMLKGIFGPPVTRVTIFGAGTIGSYT 237

Query: 245 LNILKNTNTHVKLVLELNQEQAQYFSQEFPPNVVHGDGTAKNILLEESVTSFDAVATLTG 304
Sbjct: 238 AKILAKGMTSVKLISSMERCEALSGELEGVRIVCGDATDIEFLIEEIEIGKSDAVLAATE 297

60  Query: 305 VDEENIITSMFLESIGIPKNITKVNRTSLLEIIDDKQLSSIITPKRIADVHVMHFVRGRV 364
     DE+N++ S+ +++G I KV + +++ + + + P+ + + V +R
Sbjct: 298 SDEKNLLISLLSKNLGARIAIAKVEKREYVKLFBAVGVDVALNPRSVTYNEVSKLLR--- 354

Query: 365 NAQDSNLEAMHHIANDRIETLQFEIKETSKLANRSLASLKLKONILIAAIIRNNKTIFPT 424

```

-1700-

+E + I + + + ++L ++L L L ++ +I AI+R N+ + P
 Sbjet: 355 ---TMRIETLAEIEGTAVVEV---VVRNTRLVGKALKDLPLPKDAIIGAIVRGNECLIPR 408
 Query: 425 GEDVLTGDRIVVITLLKNITRTSDM 450
 5 G+ + DR++V I + ++
 Sbjet: 409 GDTTIEYEDRLLVFAKWDEIEKIEEI 434
 Identities = 48/212 (22%), Positives = 99/212 (46%), Gaps = 15/212 (7%)
 Query: 3 VKMRIIVVGGGKVG TALCRSLVAEKHDVVLIEKKENVLKRVTKQHDIMGIV-GNGANYKI 61
 10 V R+ + G G +G+ + L V LIE + ++ + + + IV G+ + +
 Sbjet: 221 VTRRVTFGAGTIGSYTAKILAKGMTSVKLISSMERCEALSGELEGVRIVCGDATDIEF 280
 Query: 62 LEQAEVKNCDIFIAITDRDEVNMISAVLAKKMGAKETVVRMRNPEYSNPYFKDKNFLGFS 121
 15 L + E+ D +A T+ DE N++ ++L+K +GA+ + ++ EY + +G
 Sbjet: 281 LIEEEIGKSDAVLAATESDEKNLLISLLSKNLGARIAIAKVEKREYVKLF----EAVGVD 336
 Query: 122 SVVNPELLAAQYIA---NTIEFPNATSV EHFANGRVMLMEFKILEGNKLCHTSMSQIRKK 178
 +NP + ++ T+ +E A V++ +++ G L + +
 20 Sbjet: 337 VALNPRSVTYNEVSKLLRTMRIETLAEIEGTAVVEVVVRNTRLV-GKALKDLPLPK---- 391
 Query: 179 FGNIVICAIERDGLIIPDGDATIQVKDKIFV 210
 + +I AI R + +IP GD TI+ +D++ V
 Sbjet: 392 --DAIIGAIVRGNECLIPRGDTTIEYEDRLLV 421

25 There is also homology to SEQ ID 4716.

SEQ ID 8836 (GBS384) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 2; MW 53kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 6; MW 78kDa).

30 The GBS384-GST fusion product was purified (Figure 212, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 279), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1532

35 A DNA sequence (GBSx1623) was identified in *S.agalactiae* <SEQ ID 4717> which encodes the amino acid sequence <SEQ ID 4718>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4948(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1701-

Example 1533

A DNA sequence (GBSx1624) was identified in *S.galactiae* <SEQ ID 4719> which encodes the amino acid sequence <SEQ ID 4720>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -12.58    Transmembrane    37 - 53 ( 33 - 61)
    INTEGRAL    Likelihood = -11.57    Transmembrane    183 - 199 ( 179 - 214)
    INTEGRAL    Likelihood = -10.03    Transmembrane    397 - 413 ( 392 - 424)
    INTEGRAL    Likelihood = -6.79     Transmembrane    14 - 30 ( 5 - 31)
10  INTEGRAL    Likelihood = -6.42     Transmembrane    71 - 87 ( 69 - 93)
    INTEGRAL    Likelihood = -4.99     Transmembrane    278 - 294 ( 274 - 295)
    INTEGRAL    Likelihood = -4.19     Transmembrane    133 - 149 ( 132 - 152)
    INTEGRAL    Likelihood = -4.09     Transmembrane    327 - 343 ( 324 - 344)
    INTEGRAL    Likelihood = -2.44     Transmembrane    236 - 252 ( 234 - 252)
15  INTEGRAL    Likelihood = -0.59     Transmembrane    456 - 472 ( 456 - 472)

----- Final Results -----
        bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10065> which encodes amino acid sequence <SEQ ID 10066> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

25  >GP:AAB90400 GB:AE001046 TRK potassium uptake system protein (trkH)
    [Archaeoglobus fulgidus]
    Identities = 166/480 (34%), Positives = 262/480 (54%), Gaps = 10/480 (2%)

30  Query: 1  MNKSMIRFLLSKLLLEAALLAIPLTGGLTYREP-QSVMMSIVITMIILILGLGLSLFK 59
    MN + +L KLL++ + +PL ++ EP ++ +++++ +LG G +
    Sbjct: 1  MNLRLTASILGKLLMLFSFSFILPLIAAHVFEEPYHPFLIPALSLVGAVLGY-GIKTE 59

    Query: 60  PKNYHIYTKGMLIVALCWILWSFFGALPFVISGQIPNIIDAFFEVSSGFTTTGATILDD 119
    + + KE IVAL W+ S FG++P++I G P +DAFFE SGFTTTGA++L
35  Sbjct: 60  SEFDSLRLHKEFAIVALIWLWMSIFGSIPYIIFGISP--VDAFFESMSGFTTTGASVLTTP 117

    Query: 120  VSVLSPALLFWRSFTHLIGGMGVLVFALAIMENSKNSHLEVMAEVPGPVFGKVSKLKK 179
    L +LL WRS T IGGMG++V LAI N + +AE PG K+ +++
40  Sbjct: 118  EE-LPKSLLLWRSLLQWIGGMGIIVFLAIFPNVAKRSTVLFQAEYPGVSLSKLKPRI 176

    Query: 180  TAQILYLLYLLMFAVFAVILYFAGMPFFDSIIIAMGTAGTGGFAVYNDIAHYNSPLITN 239
    TA LY +YLL+ +LY G+ FD+I T TGG++ +++SIA + +
    Sbjct: 177  TALSLYKVYLLLTIAEVALLYALGLSLFDAINHTFTTLTGGSYTHSESTAFFKDVREVA 236

45  Query: 240  LVSIGMLIFGVNFNLYLLLLLRKIKAFFGDEELKTYLRIVAIATFMIALNVIGMYDNFRQ 299
    +V+ + G NF L Y LL K F + E + Y+ +A+A+ +IA + Y F +
    Sbjct: 237  VVAFFAFLGGANFALIYFLLSGK-PVIFRNTEFRAYVCFLALASVVIAAVNLDYSIF-E 294

50  Query: 300  GLEHIFFEVSAIITTTGFGVTDITRWPLFSQVILLFLMFIGGSAGSTAGGFKVMRSLILA 359
    L + F+ +I+TTTGF D W +++IL+ LMFIGGS+GST GG KV+R +L
    Sbjct: 295  SLRYSIFQAVSIMTTTGFTTADFDAWSDSAKLILVVLFMFIGGSSTGGGKIVIRIYLLI 354

    Query: 360  KIARNQVLSTLYPNRVMSLHINKSVLKDNTQHGVLKYLTIYLAIFMALVVLTLDTNDFL 419
    K A +Q+L P V ++ + K + + +Y+ IF ++++L D +
55  Sbjct: 355  KYAVHQILRAAEPRTVRAVKFEGRRAIKKEILDDIAAFFVLYILIFAVSSILVSLSGYDIV 414

    Query: 420  VVISAAASCFFNNIGP---LLGSNETFSFFSPFSKLLLSFAMIAGRLEIYPVLLMFIPKRW 476
    ISA A+ N+GP L G+ E ++ F +K+LL+ M GRLEI+ V+ +FIP W
60  Sbjct: 415  TSISATAATLGNVGPLGLAGAAENYASFPSLTKILLAVNMWIGRLEIFTVVSIFPTFW 474

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-1702-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1534

A DNA sequence (GBSx1625) was identified in *S.agalactiae* <SEQ ID 4721> which encodes the amino acid sequence <SEQ ID 4722>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence (or aa 1-20)

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2870(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAD36530 GB:AE001797 conserved hypothetical protein
 [Thermotoga maritima]
 Identities = 43/75 (57%), Positives = 57/75 (75%), Gaps = 1/75 (1%)

 Query: 2 LKSFLIFLVRFYQKNISPAPFASCRYRPTCSTYMIEAIQKHG-LKGVLMGIARILRCHPL 60
 +K LI L+RFYQ+ ISP P +CR+ PTCS Y I+A++KHG LKG +G+ RILRC+PL
 20 Sbjct: 1 MKKLLIMLIRFYQRYISPLKPPTCRFTPTCSNYFIQALEKHGLLKGTFLGLRRLRCNPL 60

 Query: 61 AHGGNDPVPDHFSLR 75
 + GG DPVP+ FS +
 25 Sbjct: 61 SKGGYDPVPEEFSFK 75

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4723> which encodes the amino acid sequence <SEQ ID 4724>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3639(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 53/78 (67%), Positives = 60/78 (75%)

40 Query: 1 MLKSFLIFLVRFYQKNISPAPFASCRYRPTCSTYMIEAIQKHGLKGVLMGIARILRCHPL 60
 M+K LI V+ YQK ISP P SCRY+PTCS YM+ AI+KHG KG+LMGIARILRCHP
 Sbjct: 1 MMKLLIVSVKAYQKYISPLSPPCRYKPTCSAYMLTAIEKHGKTKGILMGIARILRCHPF 60

 Query: 61 AHGGNDPVPDHFSLRRNK 78
 GG DPVP+ FSL RNK
 45 Sbjct: 61 VAGGVDPVPEDFSLMRNK 78

SEQ ID 4722 (GBS233) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 3; MW 35.6kDa).

50 The GBS233-GST fusion product was purified (Figure 207, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 280), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1703-

Example 1535

A DNA sequence (GBSx1626) was identified in *S. agalactiae* <SEQ ID 4725> which encodes the amino acid sequence <SEQ ID 4726>. This protein is predicted to be ribosomal large subunit pseudouridine synthase B (rluB). Analysis of this protein sequence reveals the following:

```

5   Possible site: 18
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2957(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15 >GP:BA05295 GB:AP001512 pseudouridylate synthase [Bacillus halodurans]
   Identities = 130/239 (54%), Positives = 175/239 (72%), Gaps = 2/239 (0%)

   Query: 2  RINKYIAHAGIASRRKAEELIKQGMVTINGQVVNELATQVKAG-DLVEIEGSPYNEEKV 60
             R+ K IA AGIASRRKAE+LI +G V +NGQVV EL +V D +E+EG P+ EE V
20   Sbjct: 3  RLQKVIAQAGIASRRKAEQLILEGKVKVNGQVVKELGIKVNPNQDDIEVEGVPVEKEEPV 62

   Query: 61  YLLNKPGRGVISSVSDDKGRKTVIDLLPQVKERIYPVGRLDWDTTGLLILITNDGDFDVKM 120
             Y+LL KP GVISSV DDKGRK V D L ++++R+YPVGRLD+DT+GLL+LTNDG+F + +
25   Sbjct: 63  YFLLYKPTGVISSVSDDKGRKVVDLFL-EIEQRVYPVGRLDYDTSGLLLITNDGEFANLL 121

   Query: 121 IHPRNEIDKVYLARVKGATKENLRPLTRGVVIDGKKTTPARYTTIKVDHEKNRSVVELT 180
             +HPR++I+KVY+A+VKGI T++ L+ L RGV ++ T PA+ ++ VD K ++V+LT
30   Sbjct: 122 MHPRHKIEKVYVAKVGIPTRDQLKLLARGVKLEDTGTPAPAKVKMLSVDRRKQTAIVKLT 181

   Query: 181 IHEGRNHQVKKMFQVGLLVDKLSRTQFGTLDLTGLRPGEARRLNKKEISQLHNAAINK 239
             IHEGRN QV++MFE +G V KL R QF LDL+G+ PG+ R L E+ L A+ K
   Sbjct: 182 IHEGRNRQVRRMFETIGCEVMKLRKREQFAFLDLSGMNPQDVRPLKPIEVKHLRELAVTK 240

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4727> which encodes the amino acid sequence <SEQ ID 4728>. Analysis of this protein sequence reveals the following:

```

35   Possible site: 18
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.1587(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

45   Identities = 210/239 (87%), Positives = 228/239 (94%)

   Query: 1  MRINKYIAHAGIASRRKAEELIKQGMVTINGQVVNELATQVKAGDLVEIEGSPYNEEKV 60
             MRINKYIAHAGIASRRKAEELIKQ+VT+NGQV+ +LAT VK+GD+VEIEGSPYNEEKV
   Sbjct: 9  MRINKYIAHAGIASRRKAEELIKQGLVTLNGQVITDLATTVKSGDVVEIEGSPYNEEKV 68

50   Query: 61  YLLNKPGRGVISSVSDDKGRKTVIDLLPQVKERIYPVGRLDWDTTGLLILITNDGDFDVKM 120
             YLLNKPGRG ISSVSDDKGRKTIV+DLLPQVKERIYPVGRLDWDT+G+LILITNDGDFD M
   Sbjct: 69  YLLNKPGRGAISSVSDDKGRKTVIDLLPQVKERIYPVGRLDWDTSGVLILITNDGDFDITM 128

55   Query: 121 IHPRNEIDKVYLARVKGATKENLRPLTRGVVIDGKKTTPARYTTIKVDHEKNRSVVELT 180
             IHPRNEIDKVYLARVKGATKENLRPLTRG+VIDGKKTTPARY I++V+ +K+RS+VELT
   Sbjct: 129 IHPRNEIDKVYLARVKGATKENLRPLTRGIVIDGKKTTPARYNIVRVEADKRSIVELT 188

   Query: 181 IHEGRNHQVKKMFQVGLLVDKLSRTQFGTLDLTGLRPGEARRLNKKEISQLHNAAINK 239
             IHEGRNHQVKKMFQ VGLLVDKLSRT+FGT+DL GLRPGEARRLNKKEISQLHN A K
60   Sbjct: 189 IHEGRNHQVKKMFQSVGLLVDKLSRTFRGTVDLKGRLPGEARRLNKKEISQLHNLANTK 247

```

-1704-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1536

A DNA sequence (GBSx1627) was identified in *S.galactiae* <SEQ ID 4729> which encodes the amino acid sequence <SEQ ID 4730>. Analysis of this protein sequence reveals the following:

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1476(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:BA05280 GB:AP001512 unknown conserved protein [Bacillus halodurans]
 Identities = 75/180 (41%), Positives = 107/180 (58%), Gaps = 10/180 (5%)

Query: 6 SIEALLFVAGEDGLSLRQMAELLSLTPSALIQQLEKLAKRYEEDDDSSLLLETAQTYKL 65
 +IE +LFV G++G++L ++ +LL L+ + LE+L Y D+ L + E A ++L
20 Sbjct: 9 AIEGILFVRGDEGVLEELCDLLELSTDVVYAALEELRLSYT-DEARGLRIEEVAHAFL 67

Query: 66 VTKDSYMTLLRDYAKAPINQSLSRASLEVLISIIAYKQPITRIEIDDIRGVNSSGAITRLI 125
 TK + A + + LS+A+LE L+IIAY+QPITRIE+D++RGV S AI L
25 Sbjct: 68 STKPELAPYFKKLALSTLQSGLSQALETIAIIAYRQPITRIEVDVIRGVKSEKAIQTLT 127

Query: 126 AFGLIKEAGKKEVLGRPNLYETTNYPFLDYMGINQLDDL-----IDASSIELVDEEVSLEF 179
 + LIKE G+ + GRP LY TT FLD+ G+ L +L ID SSI EE LF
30 Sbjct: 128 SRLLIKEVGRAGGTGRPIIYGTTPQFLDHFLGKSLKELPPLPEDIDESSI---GEEADLF 184

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4731> which encodes the amino acid sequence <SEQ ID 4732>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.1062(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 130/179 (72%), Positives = 159/179 (88%)

Query: 1 MTYLGSI EALLFVAGEDGLSLRQMAELLSLTPSALIQQLEKLAKRYEEDDDSSLLLETA 60
 MTYL IEALLFVAGE+GLSLR +A +LSLTP+AL QQLEKL+++YE+D SSL L+ETA
45 Sbjct: 1 MTYLSQIEALLFVAGEEGLSLRHLASMLSLTPALQQLEKLSQKYEKDQHSSLCIETA 60

Query: 61 QTYKLVTKDSYMTLLRDYAKAPINQSLSRASLEVLISIIAYKQPITRIEIDDIRGVNSSGA 120
 TY+LVTK+ + LLR YAK P+NQSLSRASLEVLIS+AYKQPITRIEIDDIRGVNSSGA
50 Sbjct: 61 NTYRLVTKEGFAELLRAYAKTPMNQSLSRASLEVLISIVAYKQPITRIEIDDIRGVNSSGA 120

Query: 121 ITRLIAFGLIKEAGKKEVLGRPNLYETTNYPFLDYMGINQLDDLIDASSIELVDEEVSLEF 179
 +++L+AF LI+EAGKK+V+GRP+LY TT+YFLDYMGIN LD+LI+ S++E DEE++LF
55 Sbjct: 121 LSKLLAFDLIREAGKKDVVGRPHLYATTDYFLDYMGINHLDLIEVSAVEPADEEIALF 179

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1705-

Example 1537

A DNA sequence (GBSx1628) was identified in *Sagalactiae* <SEQ ID 4733> which encodes the amino acid sequence <SEQ ID 4734>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1012(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14254 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
Identities = 89/246 (36%), Positives = 145/246 (58%), Gaps = 19/246 (7%)

Query: 3 IKLKDFEGPLDLLLHLVSKYEVDIYDVPIVEVIEQYLAYIATLQAMRLEVAGEYMLMASQ 62
+K+ FEGPLDLLLHL+++ E+DIYD+P+ ++ EQYL Y+ T++ + L++A EY++MA+
Sbjct: 6 VKIDTFEGPLDLLLHLINRLEIDIYDIPVAKITEQYLLYVHTMRVLELDIASEYLVMAAT 65

Query: 63 LMLIKSRNLLPK----VVESNPI-EDDPMEELLSQLEEYRRFKVLSEELANQHQERAKYF 117
L+ IKSR LLPK + E + E+DP EL+ +L EYR++K +++L + +ER K F
Sbjct: 66 LLSIKSRMLLPKQEEELFEDELLEEDPREELIEKLIERYKYKDAADLKEREERERQKSF 125

Query: 118 SKPKQEVIFEDAILLHDKSVMDLFLTFQSQMSQKQKELSNS-----QTVIEKEDYRIED 171
+KP ++ + +S L +T M+ QK L +T I ++D IE
Sbjct: 126 TKPPSDL--SEYAKEVKQSEQKLSVTIVYDMIGAFQKVLKRKKINRPMETTITRQDIPIEA 183

Query: 172 MMIVIERHFNLLKKTT--LQEVFADCQTKSEMITLFLAMLELIKHLQITVEQDSNFSQV 228
M I +LK + T ++F + K ++ FLA+LEL+K + +EQ+ NFS +
Sbjct: 184 RMNEIVH--SLKSRGTRINFMDLF-PYEQKEHLVVTFLAVLELMKNQLVLIEQEHNFSDI 240

Query: 229 ILRKEE 234

+ E

Sbjct: 241 YITGSE 246

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4735> which encodes the amino acid sequence <SEQ ID 4736>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.61 Transmembrane 199 - 215 (199 - 218)

----- Final Results -----

bacterial membrane --- Certainty=0.2444(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14254 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
Identities = 86/239 (35%), Positives = 145/239 (59%), Gaps = 15/239 (6%)

Query: 3 IKLKDFEGPLDLLLHLVSQYKVDIYEVPIVEVIEQYLNIYETLQVMKLEVAGDYMLMASQ 62
+K+ FEGPLDLLLHL+++ ++DIY++P+ ++ EQYL Y+ T++V++L++A +Y++MA+
Sbjct: 6 VKIDTFEGPLDLLLHLINRLEIDIYDIPVAKITEQYLLYVHTMRVLELDIASEYLVMAAT 65

Query: 63 LMLIKSRRLLPKVVEHI-----EEDLEQDLLEKIEEYSRFAVSQALAKQHDORAKWY 115
L+ IKSR LLPK E + EED ++L+EK+ EY ++K ++ L ++ ++R K +
Sbjct: 66 LLSIKSRMLLPKQEEELFEDELLEEDPREELIEKLIERYKYKDAADLKEREERERQKSF 125

Query: 116 SKPKQELI-FEDAILQEDK----TVMDLFLAFSNIMAARAVLKNHTVIERDDYKIEDM 170
+KP +L + + Q ++ TV D+ AF ++ K+ + + T I R D IE
Sbjct: 126 TKPPSDLSEYAKEVKQSEQKLSVTIVYDMIGAFQKVLKRKK-INRPMETTITRQDIPIEA 184

-1706-

Query: 171 MASIKORLEKENV-IRLSAIFEECQTLNEVISIFLASLELIKLVVVFVEQLSNFGAIL 228
 M I L+ I +F Q + V++ FLA LEL+K +V +EQ NF I +
 Sbjct: 185 MNEIVHSLKSRGTRINFMDLFPYEQKEHLVVT-FLAVLELMKNQLVLIEQEHNFSDIYI 242

An alignment of the GAS and GBS proteins is shown below.

Identities = 156/235 (66%), Positives = 191/235 (80%), Gaps = 2/235 (0%)

Query: 1 MDIKLKDFEGPLDLLHLVSKYEVDIYDVPIVEVIEQYLAYIATLQAMRLEVAGEYMLMA 60
 MDIKLKDFEGPLDLLHLVVS+Y+VDIY+VPIVEVIEQYL YI TLQ M+LEVAG+YMLMA
 Sbjct: 1 MDIKLKDFEGPLDLLHLVVSQYKVDIYEVPIVEVIEQYLYNIETLQVMKLEVAGDYMLMA 60

Query: 61 SQLMLIKSRNLLPKVVESNPIEDPEMELLSQLEEYRRFKVLSEELANQHGERAKYFSKP 120
 SQLMLIKSR LLPKVVE IE+D E +LL ++EY RFK +S+ LA QH +RAK++SKP
 Sbjct: 61 SQLMLIKSRLLPKVVEH--IEEDLEQDLEKIEEYSRFAVSQALAKQHDQRAKWYSKP 118

Query: 121 KQEVIFEDAILLHDKSVMDFLTFSQMMSQKQKELSNSQTVIEKEDYRIEDMMIVIERHF 180
 KQE+IFEDAIL DK+VMDLFL FS +M+ K+ L N+ TVIE++DY+IEDMM I++
 Sbjct: 119 KQELIFEDAILQEDKTVMDFLAFSNIMAARAVLKNHNTVIERDDYKIEDMMASIKQRL 178

Query: 181 NLKKKTTLQEVFADCCQTKSEMITLFLAMLELIKLVHQTVEQDSNFSQVILRKEEK 235
 + L +F +CQT +E+I++FLA LELIKLVH +VEQ SNF +ILRKE+K
 Sbjct: 179 EKENVIRLSAIFEECQTLNEVISIFLASLELIKLVHVFVEQLSNFGAILRKEKK 233

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1538

A DNA sequence (GBSx1629) was identified in *S.agalactiae* <SEQ ID 4737> which encodes the amino acid sequence <SEQ ID 4738>. This protein is predicted to be pXO1-18. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.14 Transmembrane 128 - 144 (127 - 145)

----- Final Results -----
 bacterial membrane --- Certainty=0.2657(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA05248 GB:AP001512 integrase/recombinase [Bacillus halodurans]
 Identities = 67/271 (24%), Positives = 117/271 (42%), Gaps = 35/271 (12%)

Query: 11 LKTMINDINNFISSK----LSLNSRKSYPHYDLKQFYKII-----GGHVNSEKLALY 58
 ++T+ N++ F+ +K LS N+ +SY DLKQ+ + + ++ E + Y
 Sbjct: 1 METVNNNLQQFLHFQKVERGLSNNTIQSYGRDLKQYIQYVERVEEIRSARNITRETILHY 60

Query: 59 QQSLSEFKL--TARKRKL SAVNQFLFFLYNRGTLKEFYRL-----QETEKITLAQTKSQI 111
 L E T+ R ++A+ F FL + + T+++ A T ++
 Sbjct: 61 LYHLREQGRAETSIARAVAARSFHQFLLREKLSDSPTVHVEIPKATKRLPKALTIEEV 120

Query: 112 MDLSNFYQDTPYSGRLIALLL--SLGLTPAEIANLKKADFDTFNLS-IEKSQMKRI 168
 L N Q D S R A+L L + G+ +E+ L +D + + + K +RI
 Sbjct: 121 EALLNSPQGRDPFSLRNKAMLELLYATGMRVSELIGLTLSDIHLMSGFVRCLGKGNKERI 180

Query: 169 LKLPEDLLPFLLESLEEDG-----DLVF-EHNGKPYSRQWFFNQLTDFLNEKN-E 216
 + + + + +ES +G D VF H+G+P SRQ F+ L N +
 Sbjct: 181 IPIGQ-VATEAVESYLANGRGKLMKKQSHDHVFVNHHGRPLSRQGFWMKMLQAKNVNID 239

Query: 217 QQLTAQLLREQFILKQKENGKMTLSRLIG 247

-1707-

+ LT LR F ENG + + +LG
 Sbjct: 240 KPLTPHTLRHSFATHLLENGADLRAVQEMLG 270

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4739> which encodes the amino acid
 5 sequence <SEQ ID 4740>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.90 Transmembrane 111 - 127 (110 - 127)

10 ----- Final Results -----
 bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 117/243 (48%), Positives = 167/243 (68%), Gaps = 1/243 (0%)

20 Query: 18 INNFIESKISLNSRKSYHYDLKQFYKIIGGHVNSEKLALVQOSLSEFKLTARKRKLSAV 77
 I FI SK LS NS+K+Y YDL+QF ++IG VN +KL LYQ S++ L+A+KRKLS
 Sbjct: 5 IEPFIASKALSQNSQKAYRYDLQQFCQLIGERVNQDKLLLYQNSTANLSLSAKKRKLSTA 64

25 Query: 78 NQFLFPLYNRGTLKEFYRLQETEKITLAQTK-SQIMDLNIFYQDTPYSGRLIALLLISL 136
 NQFL++LY L ++RL +T K+ + + + I++ FYQ T + G+LI+LLIL L
 Sbjct: 65 NQFLVYLYQIKYLNLSYFRLTDTMKVMRTEKQQAIIINTDIFYQKTPFVWGQLISLLILEL 124

30 Query: 137 GLTPAEIANLKKADFDTTFNLSIEKSQMKRILKLPEDLLPFLLESLEEDGDLVFEHNGK 196
 GLTP+E+A ++ A+ D F +L+++ + R+L L + L+PFL + L +FEH G
 Sbjct: 125 GLTPSEVAGIEVANLDLNFQMLTLTKKGVRLPLSQLILIPFLEQQLVGKEVYLFEHRGI 184

35 Query: 197 PYSRQWFFNQLTDFLNEKNEQQLTAQLLREQFILKQKENGKIMTELSRLGLKTPITLER 256
 P+SRQWFFN L F+ + LTAQ LREQFILK+K GK++ ELS +LGLK+P+TLE+
 Sbjct: 185 PFSRQWFFNHLKTFVRSIGYEGLTAQKLREQFILKEKLAGKSIIELSDIILGLKSPMTLEK 244

Query: 257 YYR 259
 YY+
 Sbjct: 245 YYK 247

SEQ ID 4738 (GBS383) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 40 extract is shown in Figure 68 (lane 7; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion
 product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 5; MW 57.1kDa).

The GBS383-GST fusion product was purified (Figure 212, lane 8) and used to immunise mice. The
 resulting antiserum was used for FACS (Figure 308), which confirmed that the protein is immunoaccessible
 on GBS bacteria.

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1539

A DNA sequence (GBSx1630) was identified in *S.agalactiae* <SEQ ID 4741> which encodes the amino
 acid sequence <SEQ ID 4742>. Analysis of this protein sequence reveals the following:

50 Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2465(Affirmative) < succ>

-1708-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BAB05201 GB:AP001512 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 Identities = 38/136 (27%), Positives = 73/136 (52%), Gaps = 1/136 (0%)

10 Query: 7 ESFLNHLHDHYLIPAEVDVAIFVDTHNADHVMLLLASNGFSRVPVITKEKKYVGTISISDI 66
 ++ + N L +IP E VA ++ +H +L+L +G++ +PV+ + K G IS S I
 Sbjct: 7 QNIMDNDLKELVIPFEKVAHVHLSNPLEHALLVLIKSGYTAIPVLDEHSLHGVISKSLI 66

15 Query: 67 MGYSKSGQLTDWE-MAQTDIVEMVNTKIEPINEAATLTAIMHKIVDYPFLPVISDQNDFR 125
 + + E +A + +++N +I I+ A+ + + + +PF+ ++ D F
 Sbjct: 67 LDALLGVERIEMERLAHLVVKDVMNPEIPTIHHKASFSRALKVSIAPPFICILDDGGSFL 126

20 Query: 126 GIITRKSILKAINSL 141
 GI+TR +IL IN L
 Sbjct: 127 GILTRSTILSFINRQL 142

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4743> which encodes the amino acid sequence <SEQ ID 4744>. Analysis of this protein sequence reveals the following:

25 Possible site: 47
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3539 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 119/153 (77%), Positives = 137/153 (88%)

35 Query: 1 MIAKEFESFLNHLHDHYLIPAEVDVAIFVDTHNADHVMLLLASNGFSRVPVITKEKKYVGT 60
 MIAKEFE+FL++HLD+YLIP +D+AIF+DTHNADHVMLLL SNGFSRVPVIT+EKKYVGT
 Sbjct: 1 MIAKEFETFLMSHLNLYLIPEQDLAIFIDTHNADHVMLLLVSNNGFSRVPVITREKKYVGT 60

40 Query: 61 ISISDIMGYQSKGQLTDWEMAQTDIVEMVNTKIEPINEAATLTAIMHKIVDYPFLPVISD 120
 ISISDIM YQSK QLTDWEM+QTDI EMVNTKIE I+ ++LT IMHK++D+PFLPV+
 Sbjct: 61 ISISDIMYQSKRQLTDWEMSQTDIGEMVNTKIETISITSSLTIMHKLIDFPFLPVVDR 120

45 Query: 121 QNDFRGIITRKSILKAINSLLDHFTDDEYTTTPK 153
 N F GIITRKSILKA+NSLLDHFTD+YTI K
 Sbjct: 121 ANRFVGIITRKSILKAVNSLLDHFTDDYTIKK 153

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1540

50 A DNA sequence (GBSx1631) was identified in *S.agalactiae* <SEQ ID 4745> which encodes the amino acid sequence <SEQ ID 4746>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4421 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-1709-

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BA06785 GB:AP001517 unknown conserved protein [Bacillus halodurans]
Identities = 55/169 (32%), Positives = 95/169 (55%), Gaps = 1/169 (0%)

5   Query: 5   KLVMDSHGDRIKDIKNHYLGKVDALFHNGDSELPSSDPIWEGIHVVTCNCDYDSGY 64
      KL+++SDSHG D +K + + + +VDAI H GDSELP D EG+++V GNCD+ +
      Sbjct: 2   KLLILSDSHGWSDELKAVADKHRQEVDAIHCSDSELPDRDDRAEGMNIVRGNCDFGVDF 61

10  Query: 65   PEVLVTKIDNAVIVQTHGHLHQINFTWDKDLLAQOEDADICLYGHLHRADAWKNGKTIF 124
      PE + + + + THGHL+ + + L A++ A + +GH H A +++ +F
      Sbjct: 62   PEDFIKTVGDFNVVYTHGHLNVKMSYVSLTYRAEVGAQLVCFGSHSVATSFQENGIVF 121

      Query: 125 INPGSVLQPRGPINEKLYAVVTITDSKVLVEYYTRQHPYPNLTKELSR 173
      +NPGS+ PR E+ Y + + D ++ + + R +L + R
15  Sbjct: 122 VNPGSLRLPRNR-KEQTYCLAYVRDDQIELTFLDRDGHEVTDLQRTYLR 169

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4747> which encodes the amino acid sequence <SEQ ID 4748>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
20  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3835(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 117/173 (67%), Positives = 143/173 (82%)

30  Query: 1   MAIRKLVMSDSHGDRIKDIKNHYLGKVDALFHNGDSELPSSDPIWEGIHVVTCNCDY 60
      MA + ++VMSDSHGDRIK+ IK+ YLG+VDAIFHNGDSEL SSDPIW GI+VV GNCDY
      Sbjct: 1   MASKTIIVMSDSHGDRIKQAIKDKYLGQVDALFHNGDSELNSSDPIWAGIYVVGNCNDY 60

      Query: 61   DSGYPEVLVTKIDNAVIVQTHGHLHQINFTWDKDLLAQOEDADICLYGHLHRADAWKNG 120
      D+GYP+ LVT++ I QTHGHL+ INFTWDKLD AQ+ ADICLYGHLHR AW+ G
35  Sbjct: 61   DTGYPDLVLTQLGTVTIAQTHGHLHYHINFTWDKLDYFAQEVVADICLYGHLHRPAAWQVG 120

      Query: 121  KTIFINPGSVLQPRGPINEKLYAVVTITDSKVLVEYYTRQHPYPNLTKELSR 173
      +T+F+NPGSV QPRG INEKLYA V +TD+++ V+Y+TR H+ YP+L+KE R
40  Sbjct: 121  QTLFMNPGSVTQPRGEINEKLYARVELTDTQIKVDYFTRDHKLYPSLSKEFKR 173

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1541

45 A DNA sequence (GBSx1632) was identified in *S.galactiae* <SEQ ID 4749> which encodes the amino acid sequence <SEQ ID 4750>. This protein is predicted to be HAM1 family protein. Analysis of this protein sequence reveals the following:

```

Possible site: 23
50  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1218(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB14796 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]

```

-1710-

Identities = 96/189 (50%), Positives = 130/189 (67%), Gaps = 1/189 (0%)

Query: 128 LIATHNEGKTKEFRELPGKLGLKVENLNDYPDLPEVEETGMTFEENARLKAETISKLTGK 187
 +IATHN GK KEF+E+ G V++L + E+EETG TFEENA +KAE ++K K
 Sbjct: 8 LIATHNPGKVKEFKETLEPRGYDVKSLAEIGFTTEEIEETGHTFEENAIMKAEAVAKAVNK 67

Query: 188 MVISDDSGLKVDALGGLPGVWSARFSGPDATDARNNAKLLHELAMVFDKERRSAQFHITL 247
 MVI+DDSG L +D LGG PGV+SAR++G D N K+L EL + +KE+R+A+F L
 Sbjct: 68 MVIADDSGLSIDNLGGRPGVVSARYAGEQKDDQANIEKVLSELKGI-EKEQRTARFRCAL 126

Query: 248 VVSAPNKESLVVEAEWPGYIGTEPKGENGFYDPLFIVGEGSRTAAELSAQEKNNLSHRG 307
 VS P +E+ VE GYI EP+GE GFGYDP+FIV + +T AEL++ EKN +SHR
 Sbjct: 127 AVSIPGEETKTVEGHVEGYIAEPRGEYGFYDPIFIVKDKDKTMAELTSDEKNKISHRA 186

Query: 308 QAVRKLMEV 316
 A++KL ++
 Sbjct: 187 DALKKLSKL 195

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4751> which encodes the amino acid sequence <SEQ ID 4752>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2590 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 214/325 (65%), Positives = 253/325 (77%), Gaps = 5/325 (1%)

Query: 1 MTKTIFESKTEGNWFLGSFQAFNYFTCFG-NDESYEAIQDVFHRLLSTLKVE---GLQLH 56
 M++ I+E K E NWF+G N + +G + + I D+ + +TL E G +
 Sbjct: 14 MSEKIYBYKDNNWFIGKMTGHNLSGWGVKHTTIKKIDDLGIAATLDWENPKGYDVS 73

Query: 57 VVQMTSDFQLLAFVLVDMINQBYSRHIKVTHQKGAIVSEDDQLFLVHLPKEGTSLEKFFD 116
 VV+ S L+ F++DMINQE R IKVT H G IL+ E+ +L V+LP+ G S FF
 Sbjct: 74 VVRHQSPSLITFIIDMINQBTQREIKVTPHAGTILLMENAKLLAVYLPEGGVSTATFF- 132

Query: 117 LKNDNNFGDTILIATHNEGKTKEFRELPGKLGLKVENLNDYPDLPEVEETGMTFEENARL 176
 ++ FGD ILIAT NEGKTKEFR LFG+LG +VENLNDYP+LPEV ETG TFEENARL
 Sbjct: 133 ATSEQFGDIILIIATRNEGKTKEFRNLFGQLGYRVENLNDYPELPEVAETGTTFEENARL 192

Query: 177 KAETISKLTGKMVISDDSGLKVDALGGLPGVWSARFSGPDATDARNNAKLLHELAMVFDK 236
 KAETIS+LTGKMV++DDSGLKVDALGGLPGVWSARFSGPDATDA+NNAKLLHELAMVFD+
 Sbjct: 193 KAETISRLTGKMLADDSGLKVDALGGLPGVWSARFSGPDATDAKNNAKLLHELAMVFDQ 252

Query: 237 ERRSAQFHITLVVSAPNKESLVVEAEWPGYIGTEPKGENGFYDPLFIVGEGSRTAAELS 296
 ++RSAQFHITLVV+APNK+SLVVEA+WPGYI T+PKGENGFYDPLFIVGE ADEL
 Sbjct: 253 KKRSAQFHITLVVAAPNKDSLVEADWPGYIATQPKGENGFYDPLFIVGETGHHAEELE 312

Query: 297 AQEKNNLSHRGQAVRKLMEVFPKWQ 321
 A +KN LSHRGQAVRKLMEVFP WQ
 Sbjct: 313 ADQKNQLSHRGQAVRKLMEVFPWQ 337

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1711-

Example 1542

A DNA sequence (GBSx1633) was identified in *S.agalactiae* <SEQ ID 4753> which encodes the amino acid sequence <SEQ ID 4754>. This protein is predicted to be glutamate racemase (murf). Analysis of this protein sequence reveals the following:

```

5   Possible site: 45
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.86    Transmembrane  114 - 130 ( 114 - 130)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10067> which encodes amino acid sequence <SEQ ID 10068> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAF72713 GB:AF263927 glutamate racemase [Carnobacterium sp. St2]
      Identities = 160/267 (59%), Positives = 202/267 (74%), Gaps = 3/267 (1%)

20  Query: 27  MDSRPIGFLDSGVGGLTVVKEFRQLPEEEVIFIGDQARAPYGP RPAAQQIREFTWQM VNF 86
      M + IGF+DSGVGGLTVVKE RQLP E + ++GD AR PYGPRP Q+R+FTW+M +F
      Sbjct: 1  MKKQAIGFIDSGVGGLTVVKEAMRQLPNESIYYVGD TARCPYGP RPEDQVRKFTWEMTHF 60

25  Query: 87  LLTKNVKMI VIACNTATATAVAMQEIKEKLDIPVLGVILPGASAAIKSTNLGKVGIIIGTPMT 146
      LL KN+KM+VIACNTATA A ++IK+KL IPV+GVILPG+ AAIK+T+ ++G+IGT T
      Sbjct: 61  LLDKNIKMLVIACNTATAAALKDIKKLAIPVIGVILPGSRAAIKATHTNRIGVIGTEGT 120

30  Query: 147 VKSDAYRQKIQALSPNTAVVSLACPKFVPIVESNQSSSLAKKVYETLSPLVGK-LDTL 205
      VKS+ Y++ I + V SLACPKFVP+VESN+ SS++AKKV ETL PL + LDTL
      Sbjct: 121 VKSNQYKKMIHSDTKALVTSIACPKFVPLVESNEYSSAIKKVVAETLRPLKNEGLDTL 180

35  Query: 206 ILGCTHYPLLRPIIQNVMGAEVKLIDSGAETVRDISVLLNYFEINHNWQNKH-GGHHFYT 264
      ILGCTHYPLLRPIIQN +G V LIDSGAETV ++S +L+YF + + QNK +FYT
      Sbjct: 181 ILGCTHYPLLRPIIQNTLGDSVTLIDSGAETVSEVSTILDYFNLAVDSONKEKAERNFYT 240

    Query: 265 TASPKGFKIEAEQWLS-QEINVERIVL 290
      T S + F IA +WL ++ VE I L
      Sbjct: 241 TGSSQMFHAIASEWLQLDDLA VEHITL 267

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4755> which encodes the amino acid sequence <SEQ ID 4756>. Analysis of this protein sequence reveals the following:

```

    Possible site: 19
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -1.70    Transmembrane  88 - 104 ( 86 - 104)

45  ----- Final Results -----
      bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50

```

The protein has homology with the following sequences in the databases:

```

    >GP:AAF72713 GB:AF263927 glutamate racemase [Carnobacterium sp. St2]
      Identities = 149/267 (55%), Positives = 202/267 (74%), Gaps = 3/267 (1%)

55  Query: 1  MDTRPIGFLDSGVGGLTVVCELI RQLPHEKIVYIGDSARAPYGP RPKKQIKEYTWELVNF 60
      M + IGF+DSGVGGLTVV E +RQLP+E I Y+GD+AR PYGPRP+ Q++++TWE+ +F
      Sbjct: 1  MKKQAIGFIDSGVGGLTVVKEAMRQLPNESIYYVGD TARCPYGP RPEDQVRKFTWEMTHF 60

    Query: 61  LLTONVKMIVFACNTATATAVAAEEVKAALDIPVLGVVLP GASAAIKSTTKGVGVIGTPMT 120

```

-1712-

LL +N+KM+V ACNTATA A +++K L IPV+GV+LPG+ AAIK+T ++GVIGT T
 Sbjct: 61 LLDKNIKMLVIACNTATAAALKDIKKKLAIPVIGVILPGSRAAIKATHNTRIGVIGTEGT 120
 Query: 121 VASDIYRKKIQLLAPSIQVRSIACPKFVPVIVESNEMCSSIAKKIVYDSLAPLVGK-IDTL 179
 5 V S+ Y+K I V SLACPKFVP+VESNE S+IAKK+V ++L PL + +DTL
 Sbjct: 121 VKSNQYKKMIHSDTKALVTSLACPKFVPLVESNEYSSAIKKVVAETLRPLKNEGLDTL 180
 Query: 180 VLGCTHYPLLRPIIQNVMGPSVKLIDSGAECVRDISVLLNYFDIN-GNYHQKAVEHRFFT 238
 +LGCTHYPLLRPIIQN +G SV LIDSGAE V ++S +L+YF++ + +++ E F+T
 10 Sbjct: 181 ILGCTHYPLLRPIIQNTLGDSVTLIDSGAETVSEVSTILDYFNLAVDSONKEKAERNFYT 240
 Query: 239 TANPEIFQEIASIWLK-QKINVEHVTL 264
 T + ++F IAS WL+ + VEH+TL
 Sbjct: 241 TGSSQMFHAIASEWLQLDDLAVEHITL 267
 15

An alignment of the GAS and GBS proteins is shown below.

Identities = 195/264 (73%), Positives = 231/264 (86%)
 Query: 27 MDSRPIGFLDSGVGGLTVVKEMFRQLPEEEVIFIGDQARAPYGPRAQQIREFTWQMVNF 86
 20 MD+RPIGFLDSGVGGLTVV E+ RQLP E++++IGD ARAPYGP RP +QI+E+TW++VNF
 Sbjct: 1 MDTRPIGFLDSGVGGLTVVCELIRQLPHEKIVYIGDSARAPYGP RP KKQIKEYTWELVNF 60
 Query: 87 LLTKNVKMIVACNTATAVAWQEIKEKLDIPVLGVILPGASAAIKSTNLGKVGIIIGTPT 146
 25 LLT+NVKMIV ACNTATAVAW+E+K LDIPVLGV+LPGASAAIKST G+VG+IGTPT
 Sbjct: 61 LLTQNVKMIVFACNTATAVAWEEVKAALDIPVLGVILPGASAAIKSTTTKGQVGVIGTPT 120
 Query: 147 VKSDAYRQKIQALSPTAVVSLACPKFVPVIVESNQMSSSLAKKVYETLSPLVGKLDTLI 206
 V SD YR+KIQ L+P+ V SLACPKFVPVIVESN+M SS+AKK+VY++L+PLVGK+DTL+
 30 Sbjct: 121 VASDIYRKKIQLLAPSIQVRSIACPKFVPVIVESNEMCSSIAKKIVYDSLAPLVGKIDTLV 180
 Query: 207 LGCTHYPLLRPIIQNVMGAEVKLIDSGAETVRDISVLLNYFEINHWNQKHGGHHFYTTA 266
 LGCTHYPLLRPIIQNVG VKLIDSGAE VRDISVLLNYF+IN N+ K H F+TTA
 Sbjct: 181 LGCTHYPLLRPIIQNVMGPSVKLIDSGAECVRDISVLLNYFDINGNYHQKAVEHRFFTTA 240
 35 Query: 267 SPKGFKEIAEQWLSQEINVERIVL 290
 +P+ F+EIA WL Q+INVE + L
 Sbjct: 241 NPEIFQEIASIWLKQKINVEHVTL 264

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 40 vaccines or diagnostics.

Example 1543

A DNA sequence (GBSx1634) was identified in *S. galactiae* <SEQ ID 4757> which encodes the amino
 acid sequence <SEQ ID 4758>. Analysis of this protein sequence reveals the following:

Possible site: 21
 45 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -11.36 Transmembrane 3 - 19 (1 - 27)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5543(Affirmative) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13675 GB:Z99113 alternate gene name: yoxG [Bacillus subtilis]
 55 Identities = 26/72 (36%), Positives = 42/72 (58%)
 Query: 1 MSITIWILLIIVALFGGLVGGIFIAKQIEKEIGEHPRLTPDAIREMMSQMGKPSKAV 60
 M++ + IL+ +VAL G+ G FIARK + + ++P + +R MM QMG KPS+ K+
 Sbjct: 1 MTLWVGILVGVVALLIGVALGFFIARKYMSYLLKKNPPINEQMLRMMMMQMGKPSQKKI 60
 60 Query: 61 QQTYRNIVKHAK 72

-1713-

Q + + K
 Sbjct: 61 NQMMKAMNNQTK 72

5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4759> which encodes the amino acid sequence <SEQ ID 4760>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -10.72 Transmembrane 7 - 23 (1 - 27)
 10 ----- Final Results -----
 bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 62/79 (78%), Positives = 69/79 (86%)
 Query: 1 MSITIWILLIIVALFGGLVGGIFIAKQIEKEIGEHPRLTPDAIREMMSQMGQKPSEAKV 60
 MS IWILL+IVAL G+ GGIFIARKQIEKEIGEHPRLTP+AIEMMSQMGQKPSEAK+
 20 Sbjct: 1 MSTAIWILLIIVALGVGVFGGIFIAKQIEKEIGEHPRLTPEAIREMMSQMGQKPSEAKI 60
 Query: 61 QQTYRNIVKHAKTAIKTKK 79
 QQTYRNI+K +K A+ K
 25 Sbjct: 61 QQTYRNIKQSKAAVSKGK 79

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1544

30 A DNA sequence (GBSx1635) was identified in *S.galactiae* <SEQ ID 4761> which encodes the amino acid sequence <SEQ ID 4762>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -7.86 Transmembrane 82 - 98 (79 - 103)
 35 ----- Final Results -----
 bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1545

45 A DNA sequence (GBSx1636) was identified in *S.galactiae* <SEQ ID 4763> which encodes the amino acid sequence <SEQ ID 4764>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence
 50 INTEGRAL Likelihood = -10.83 Transmembrane 56 - 72 (50 - 105)
 INTEGRAL Likelihood = -7.27 Transmembrane 27 - 43 (17 - 48)
 INTEGRAL Likelihood = -6.26 Transmembrane 76 - 92 (73 - 105)
 INTEGRAL Likelihood = -4.83 Transmembrane 119 - 135 (118 - 141)
 INTEGRAL Likelihood = -1.65 Transmembrane 160 - 176 (160 - 176)

-1714-

----- Final Results -----

5 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8837> which encodes amino acid sequence <SEQ ID 8838> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4765> which encodes the amino acid sequence <SEQ ID 4766>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have a cleavable N-term signal seq.

15 INTEGRAL Likelihood = -10.99 Transmembrane 45 - 61 (37 - 94)
 INTEGRAL Likelihood = -7.06 Transmembrane 74 - 90 (62 - 94)
 INTEGRAL Likelihood = -3.45 Transmembrane 110 - 126 (108 - 130)
 INTEGRAL Likelihood = -2.18 Transmembrane 149 - 165 (149 - 165)
 INTEGRAL Likelihood = -1.91 Transmembrane 21 - 37 (20 - 37)

20 ----- Final Results -----

 bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 111/173 (64%), Positives = 145/173 (83%)

30 Query: 12 MSKKTQTMVSYTSILVAFAIMIPIIMPAKIIIGPASFTLASHVPLFLSIFISVPVAILVA 71
 M+KK TQ+++YTSILVAFAI+IPIIMP K+IIGPASFTLASHVPLFL+IF+S+PVAILVA
 Sbjct: 1 MTKKPTQLIAYTSILVAFAILIPIIMPLKLIIGPASFTLASHVPLFLAIFMSIPVAILVA 60

 Query: 72 LGTGLGFLLAGFPFIVIVLRALSHIGFALIAAFLIKSKPSLLMSKWQTLLEFAVAINIIHGL 131
 LGT LGFLLAG P++IVLRALSH+ FA++AA+ + KP L+ S + FA IN+IHGL
 35 Sbjct: 61 LGTTLGFLLAGPLIIVLRALSHLLFAILAAWWSRKPQLMTSAVKCFSAFFFINVIHGL 120

 Query: 132 LEFITYVYIIITMTSNSSSTYLWSLFSLIGLGSLLHGLVDFYIALFIWKWMTQKL 184
 EF+ VYI+T T+ +S +Y WS+ LIGLGSLLHG++DFY+AL +W+++ + L
 40 Sbjct: 121 AEFLVVYIILTATTATSMYSFWSMLGLIGLGSLLHGILDFYLAIVLWRFLAKNL 173

A related GBS gene <SEQ ID 10789> and protein <SEQ ID 10790> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 3

SRCPLG: 0

45 McG: Length of UR: 24
 Peak Value of UR: 3.16
 Net Charge of CR: 2

McG: Discrim Score: 12.56

GvH: Signal Score (-7.5): -0.16

50 Possible site: 19

>>> Seems to have a cleavable N-term signal seq.

Amino Acid Composition: calculated from 20

ALOM program count: 5 value: -10.83 threshold: 0.0

55 INTEGRAL Likelihood = -10.83 Transmembrane 45 - 61 (39 - 94)
 INTEGRAL Likelihood = -6.26 Transmembrane 65 - 81 (62 - 94)
 INTEGRAL Likelihood = -4.83 Transmembrane 108 - 124 (107 - 130)
 INTEGRAL Likelihood = -1.65 Transmembrane 149 - 165 (149 - 165)
 INTEGRAL Likelihood = -0.27 Transmembrane 24 - 40 (24 - 40)
 PERIPHERAL Likelihood = 0.42 86

-1715-

modified ALOM score: 2.67
icml HYPID: 7 CFP: 0.533

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1546

A DNA sequence (GBSx1637) was identified in *S.agalactiae* <SEQ ID 4767> which encodes the amino acid sequence <SEQ ID 4768>. This protein is predicted to be transcriptional regulator, biotin repressor family. Analysis of this protein sequence reveals the following:

Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2237(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14749 GB:Z99118 yrxA [Bacillus subtilis]
Identities = 72/165 (43%), Positives = 112/165 (67%), Gaps = 2/165 (1%)

Query: 6 RRENILTTLKGTKEAISASTLAKIFSRSQVIVGDIALLRQQCDIISTPKGYL-MSSAL 64
RR+ +L LK +K ++ LAK +VSRQVIV DI+LL+A+ II+T +GY+ M +A
Sbjct: 12 RRDQLLLWLKESKSPITGGELAKKANVSRQVIVQDISLLKAKNPPIIATSQGVYMDAAA 71

Query: 65 STHQFTARLV-COHGIEQTEEBEELIILRYQGIIMNVEVEHPHYGMLTAPLNIQSQKDIDN 123
HQ R++ C HG E+TEEBL++I+ + +V++EHP+YG LTA + + ++K++ +
Sbjct: 72 QQHQQAERIIACLHGPERTEEBELQLIVDEGVTVKDVKIEHPVYGDLTAAIQVGTREKVVSH 131

Query: 124 FTAKLKVSNAELSSSLTDGLHMHMISCQDQSVFDQICEALKKAGI 168
F K+ +NA LS LTDG+H H ++ D+ DQ C+AL++AGI
Sbjct: 132 FIKKINSTNAAYLSQLTDGVHLHTLTAPDEHRIDQACQALEEAGI 176

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4769> which encodes the amino acid sequence <SEQ ID 4770>. Analysis of this protein sequence reveals the following:

Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2971(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 109/170 (64%), Positives = 136/170 (79%)

Query: 1 MKAQERRENILTTLKGTKEAISASTLAKIFSRSQVIVGDIALLRQQCDIISTPKGYLM 60
MKA++RR+ I+ L ++A+SA+ L K+ VSRQVIVGDIALLRQQ DIISTPKGY+M
Sbjct: 1 MKAEDRRQKIIECLNSEQKAVSATRLGKLLGVSRQVIVGDIALLRQQIDIISTPKGYIM 60
Query: 61 SSALSTHQFTARLVCOHGIEQTEEBEELIILRYQGIIMNVEVEHPHYGMLTAPLNIQSQKD 120

-1716-

S+AL +HQF AR+VCQH +E+T++ELEIIL +QGII VEVEHPIYGM+TAPLNI++ D
 Sbjct: 61 STALYSHQFCARIVCQHNVEETKKELEIILAHQGIITTVVEHPIYGMITAPLNIKTHSD 120

Query: 121 IDNFTAKLKVSNAELLSSLTDLGLHMHMISCQDQSVFDQICEALKKAGILY 170
 + NF +KL S AELLSSLT+GLH+H+ISC Q F I L+ AGILY
 Sbjct: 121 VTNFMKLSQSKAELLSSLTGLHSHLISCPSEAFLEIKHDLLELAGILY 170

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 1547

A DNA sequence (GBSx1638) was identified in *S.agalactiae* <SEQ ID 4771> which encodes the amino acid sequence <SEQ ID 4772>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

15	INTEGRAL	Likelihood = -8.44	Transmembrane	143 - 159 (138 - 165)
	INTEGRAL	Likelihood = -8.17	Transmembrane	164 - 180 (160 - 184)
	INTEGRAL	Likelihood = -7.17	Transmembrane	56 - 72 (53 - 78)
	INTEGRAL	Likelihood = -5.63	Transmembrane	24 - 40 (21 - 44)
	INTEGRAL	Likelihood = -4.94	Transmembrane	113 - 129 (108 - 131)
20	INTEGRAL	Likelihood = -2.39	Transmembrane	86 - 102 (86 - 103)
	INTEGRAL	Likelihood = -1.06	Transmembrane	203 - 219 (203 - 219)

----- Final Results -----

25	bacterial membrane	--- Certainty=0.4376 (Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000 (Not Clear) < succ>
	bacterial cytoplasm	--- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10069> which encodes amino acid sequence <SEQ ID 10070> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC18360 GB:AF064763 putative membrane spanning protein
 [Lactococcus lactis subsp. cremoris]
 Identities = 97/188 (51%), Positives = 133/188 (70%)

35 Query: 38 IMLYMFQNMIAIMQKMPGLYFGAIIELVLVVFVSGAARRNTPAALPLFLIYSALNGFT 97
 IM+ F NM AI+Q I+ LV+V G A +N+ ALP+F+ Y+A GF
 Sbjct: 1 IMITFFLDNMRAILQSGSLFLLVLWIIPLVMVVSLOGLAMKNSKMLPFI FIGYAAFMGFL 60

40 Query: 98 LSFIIARYTQTTVLQAFITSAAVFFAMALIGAKTKDLSCMRKALMAALIGILIASLVNL 157
 +SF + YT T + AFIT++A+FF +++ G TK++LSGM KAL A+ G+++A L+NL
 Sbjct: 61 ISFTLLMYTATDITLAFITASAMFFGLSVYGRFTRKRLSGMGKALGVAVWGLIVAMLNL 120

Query: 158 FIGSGGMSYIISIVCVIIFSGLIAIDNQMIKYVYNSQGGQVADGWAVSMALSLYLDFINL 217
 F S G++ +IS+V V+IFSGLIA+DNQ I VYN+ GQV+DGWA+SMALSLYLDFIN+
 45 Sbjct: 121 FFASTGLTILISLVGVVIFSGLIAWDNQKITQVYNAHNGQVSDGWAISMALSLYLDFINM 180

Query: 218 FLNLRFL 225
 FL +LRLF
 Sbjct: 181 FLFLRLFL 188

50

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4773> which encodes the amino acid sequence <SEQ ID 4774>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

55	INTEGRAL	Likelihood = -8.97	Transmembrane	143 - 159 (138 - 165)
	INTEGRAL	Likelihood = -5.89	Transmembrane	164 - 180 (160 - 184)
	INTEGRAL	Likelihood = -5.68	Transmembrane	56 - 72 (55 - 77)
	INTEGRAL	Likelihood = -4.78	Transmembrane	113 - 129 (110 - 130)

-1717-

INTEGRAL Likelihood = -2.81 Transmembrane 203 - 219 (203 - 222)
 INTEGRAL Likelihood = -2.76 Transmembrane 24 - 40 (23 - 41)
 INTEGRAL Likelihood = -2.76 Transmembrane 86 - 102 (86 - 104)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:AAI18360 GB:AF064763 putative membrane spanning protein
 [Lactococcus lactis subsp. cremoris]
 Identities = 90/189 (47%), Positives = 133/189 (69%)

15 Query: 38 LMLYPFRENILSILVNQPMIYYGAAIIEILVVFVASSAARKNTPAALPIFLIYSALNGFT 97
 +M+ F +N+ +IL + + II L++V A KN+ ALPIF+ Y+A GF
 Sbjct: 1 IMITFFLDNMRAILQSGSLFLLVLWIIPLVMVSLQGLAMKNSKMLPIFIGYAAFMGFL 60

20 Query: 98 LSFIIIVAYQTTVFQAFLLSSAAVFFAMSIIGVTKRDMSGLRKAMFAALIGVVVASLINL 157
 +SF ++ Y T + AF++++A+FF +S+ G TKR++SG+ KA+ A+ G++VA L+NL
 Sbjct: 61 ISFTLLMYTATDITLAFITASAMFFGLSVYGRFTKRNLSGMGKALGVAVWGLIVAMLLNL 120

25 Query: 158 FIGSGMMSYVISVISVLIFSGLIASDNQMIKRVYQATNGQVGDGWAVAMALSLYLDFINL 217
 F S ++ +IS++ V+IFSGLIA DNQ I +VY A NGQV DGWA++MALSLYLDFIN+
 Sbjct: 121 FFASTGTILISILVGVVIFSGLIAWDNQKITQVYNAHNGQVSDGWAISMALSLYLDFINM 180

30 Query: 218 FISLLRIFG 226
 F+ LLR+FG
 Sbjct: 181 FLFLRLRFG 189

An alignment of the GAS and GBS proteins is shown below.

Identities = 167/229 (72%), Positives = 202/229 (87%)

35 Query: 1 MNDNVIYTQSDSGLNQFFAKIYGLVGIGVGLSAAVSAIMLYMFPQNMIAIMQKMPGLYFG 60
 MND+VIYTQSD GLNQFFAKIY LVG+GVGLSA VS +MLY F +N+I+I+ P +Y+G
 Sbjct: 1 MNDHVIYTQSDVGLNQFFAKIYSLVGMGVGLSAFVSYLMLYPFRENILSILVNQPMIYYG 60

40 Query: 61 AIIIELVLVFVASGAARRNTPAALPLFLIYSALNGFTLSFIIARYTQTTLVQAFITSAAV 120
 A I+EL+LVFVAS AAR+NTPAALP+FLIYSALNGFTLSFII Y QTTV QAF++SAAV
 Sbjct: 61 AAIIEILVLFVASSAARKNTPAALPIFLIYSALNGFTLSFIIIVAYQTTVFQAFLLSSAAV 120

45 Query: 121 FFAMALIGAKTKKDLSCMRKALMAALIGILIASLVNLFISGGMSYIISIVCVIIFSGLI 180
 FFAM++IG KTK+D+SG+RKA+ AALIG+++ASL+NLFIGSG MSY+IS++ V+IFSGLI
 Sbjct: 121 FFAMSIIGVTKRDMSGLRKAMFAALIGVVVASLINLFIGSGMMSYVISVISVLIFSGLI 180

50 Query: 181 AYDNQMIKVVYNSQGGQVADGWAVSMALSLYLDFINLFLNILRLRFARND 229
 A DNQMIK VY + GQV DGWAV+MALSLYLDFINLF+++LR+F RND
 Sbjct: 181 ASDNQMIRVYQATNGQVGDGWAVAMALSLYLDFINLFLISLLRIFGRND 229

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1548

A DNA sequence (GBSx1639) was identified in *S.galactiae* <SEQ ID 4775> which encodes the amino acid sequence <SEQ ID 4776>. Analysis of this protein sequence reveals the following:

55 Possible site: 44
 >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2495(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1718-

A related GBS nucleic acid sequence <SEQ ID 10071> which encodes amino acid sequence <SEQ ID 10072> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4777> which encodes the amino acid sequence <SEQ ID 4778>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3277(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 127/163 (77%), Positives = 141/163 (85%)

Query: 7 YQDDKDFMDLVGHLIDHPRFQKLEAIVQHHHSTRLEHSINVSYTSYKIAKKFGWDASSTA 66
Y +DK++M+ VGHLI HPRFQKL IVQH HSTRLEHSINVSY+SYK+AK+FGWDA STA
20 Sbjct: 3 YTEDKEYMEHVCHLIAHPRFQKLSHIVQHQHSTRLEHSINVSYSSYKLAKRFGWDAKSTA 62
Query: 67 RGGLLHDDFFYYDWRVTKFNKSHAWVHPRIAVRNARKLTDLNAREEDIILKHMWGATIAPP 126
RGGLLHDDFFYYDWRVTKFNK HAWVHPRIAVRNA+KLT+LN +EEDIILKHMWGATIA P
25 Sbjct: 63 RGGLLHDDFFYYDWRVTKFNGHAWVHPRIAVRNAKKLTENKKEEDIILKHMWGATIAPP 122
Query: 127 RYKESYIVTMVDKYWAVREASRPLKRIFKKPIRFSRKFLGSHN 169
RYKESYIVTMVDKYWAV+EA PL++ + RK L SHN
Sbjct: 123 RYKESYIVTMVDKYWAVKEAVTPLRQKWSNRRFLRRKTLQSHN 165

- 30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1549

A DNA sequence (GBSx1640) was identified in *S.agalactiae* <SEQ ID 4779> which encodes the amino acid sequence <SEQ ID 4780>. Analysis of this protein sequence reveals the following:

35 possible site: 37
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -3.03 Transmembrane 213 - 229 (212 - 229)
40 ----- Final Results -----
bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 45 A related GBS nucleic acid sequence <SEQ ID 9413> which encodes amino acid sequence <SEQ ID 9414> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14825 GB:Z99118 similar to rRNA methylase [Bacillus subtilis]
Identities = 96/228 (42%), Positives = 143/228 (62%), Gaps = 5/228 (2%)

50 Query: 3 QKKYRKSSYLIEGWHLFEEAEKYGAQFLNIFVT-ETAIDR-LRKPERAIVTDDVLKELT 60
+++ + +++LIEG HL EEA K I V ET I L + +++D +T
Sbjct: 22 KERTKTNTFLIEGEHLVEEALKSPGIVKEILVKDETRIPSDLETGIQCYMLSEDAFSAVT 81
Query: 61 DSQTPQGIVAEIAFQETRWTDIKKGRFLVLEDDVQDPGNLGTMTVRTADAANFDAVFLSQKS 120

-1719-

+++TPQ I A E + +K L+++ VQDPGNLGT+RTADAA DAV L +
 Sbjct: 82 ETETPQQIAAVCHMPEEKLATARK--VLLIDAVQDPGNLGT+RTADAAAGLDVAVLGDGT 139

Query: 121 ADLYNQKTLRSMQGSFHLPVFRVEIEQFVNFKAEGITMIATTLSEQSVNYKNLPKYDY 180
 AD +N KTLRS QGSFHH+PV R + +V+ KAEG+ + T L + Y+ +P+ +
 Sbjct: 140 ADAFNGKTLRSAQGSFHIPVVRRLPSYVDELKAEGVKVYGITAL-QNGAPYQEIPQSES 198

Query: 181 FALIMGNEQGQISKMTTEADVLAHIEMPGQAESLNVAAGVVIFSL 228
 FALI+GNEG G+ + E+ D+ ++ + GOAESLNVAAGVAAAA L
 Sbjct: 199 FALIVGNEGAGVDAALLEKTDNLNLYVPLYGQAESLNVAAGVAAAA L 246

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4781> which encodes the amino acid sequence <SEQ ID 4782>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.97 Transmembrane 229 - 245 (228 - 245)

----- Final Results -----
 bacterial membrane --- Certainty=0.2190(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 141/229 (61%), Positives = 178/229 (77%)

Query: 1 MLQKKYRKSSYLIEGWHLFEEAEKYGAQFLNIFVTETAIIDLRLKPERAIVVTDVVKELT 60
 +LQKK+RK SYLIEGWHLFEEA+K G F +IFV E ++RL + ++V+ VLKELT
 Sbjct: 17 LLQKKHRKQSYLIEGWHLFEEAQKSGQVFRHIFVLEEMVERLAGEQELVIVSPQVLKELT 76

Query: 61 DSQTPQGIIVAEIAFQETRWTDIKGRFLVLEDVQDPGNLGT+RTADAA FD VFLS+KS 120
 DS +PQGIIVAE+ + + KG++LVLEDVQDPGNLGT++RTADAA FD VFLS+KS
 Sbjct: 77 DSPSPQGIIVAEVEIPKLAPPSDYKGKYLLEDVQDPGNLGT+RTADAA RFDGVFLSEKS 136

Query: 121 ADLYNQKTLRSMQGSFHLPVFRVEIEQFVNFKAEGITMIATTLSEQSVNYKNLPKYDY 180
 AD+YNQKTLRSMQGSFHLPVFR++ Q + ++ATTL++SV+YK+L ++
 Sbjct: 137 ADIYNQKTLRSMQGSFHLPVFRVTDVYQLCRELQEYETPILATTLSSKKSVDYKSLTHHER 196

Query: 181 FALIMGNEQGQISKMTTEADVLAHIEMPGQAESLNVAAGVVIFSLI 229
 AL++GNEGQGIS M AD L HI MPGQAESLNVAAG++IFSLI
 Sbjct: 197 LALVLGNEGQGISAEAAALADQLVHITMPGQAESLNVAAGVILIFSLI 245

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8839> and protein <SEQ ID 8840> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7
 McG: Discrim Score: -7.98
 GvH: Signal Score (-7.5): -3.86
 Possible site: 37
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -3.03 threshold: 0.0
 INTEGRAL Likelihood = -3.03 Transmembrane 213 - 229 (212 - 229)
 PERIPHERAL Likelihood = 5.14 149
 modified ALOM score: 1.11

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1721-

Sbjct: 1 MKALKIRVEGIVQGVGFRYFTRRVAKSLG-VKGYVMNMDGVSFIHAEG-DENALRRFLN 58

Query: 84 KIRKGPSKWSKVTVYVDIKLDNFDDFND 111

++ KGP + VT V ++ + + DF

5 Sbjct: 59 EVAKGPPA-AVVTNVSVETTPEGYEDF 85

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4785> which encodes the amino acid sequence <SEQ ID 4786>. Analysis of this protein sequence reveals the following:

Possible site: 34

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2433(Affirmative) < succ>

15 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 69/95 (72%), Positives = 85/95 (88%)

20 Query: 19 KRGQVMKKVHLIVSGRVQGVGFRYATYSLALEIGDIYGRVWNNDDGTVEILAQSTDNSKM 78

K +M+KV LIVSGRVQGVGFRYAT++LAL+IGDIYGRVWNN+DGTVEILAQSD+K+

Sbjct: 7 KEALLMQKVRLLIVSGRVQGVGFRYATHLTALDIGDIYGRVWNNDDGTVEILAQSKDSDKI 66

Query: 79 TQFIQKIRKGPSKWSKVTVYVDIKLDNFDDFNDFKM 113

25 FIQ++RKGPSKW+KVTVYVD+ + NF+DF DF++

Sbjct: 67 ATFIQEVRRKGPSKWAKVTVYVDVTMANFEDFQDFQI 101

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 1551

A DNA sequence (GBSx1642) was identified in *S.agalactiae* <SEQ ID 4787> which encodes the amino acid sequence <SEQ ID 4788>. This protein is predicted to be membrane protein homolog (yidC). Analysis of this protein sequence reveals the following:

Possible site: 16

35 >>> May be a lipoprotein

INTEGRAL Likelihood = -12.52 Transmembrane 60 - 76 (54 - 83)

INTEGRAL Likelihood = -3.66 Transmembrane 178 - 194 (177 - 196)

INTEGRAL Likelihood = -2.76 Transmembrane 140 - 156 (137 - 157)

40 INTEGRAL Likelihood = -2.60 Transmembrane 216 - 232 (213 - 232)

----- Final Results -----

bacterial membrane --- Certainty=0.6010(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10075> which encodes amino acid sequence <SEQ ID 10076> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF03934 GB:AF139908 membrane protein homolog [Listeria
monocytogenes]

50 Identities = 82/222 (36%), Positives = 133/222 (58%), Gaps = 4/222 (1%)

Query: 44 PMANLITYFAQHQLGLGFGVAIIIVTVIVRVVILPLGLYQSWKASYQAEKMAFYKPLFEPI 103

P + I + A+ G +G+AIIII T+++R +I+PL L + KMA KP + I

55 Sbjct: 3 PFTSFIMFVAKFVGGNYGIAIIITLLIRALIMPLNLRITAKAQMGMSKMAVAKPEIDEI 62

-1722-

Query: 104 NERLRNAKTQEEKLAAQTELMTAQRENGLSMFGGIGCLPLLIQMPFFSAIFFAARYTPGV 163
 RL+ A ++EE+ Q E+M + ++ +GCLPLLIQMP A ++A R + +
 Sbjct: 63 QARLKRATSKKEEQATIQKEMMAVYSKYNNINPMQ-MGCLPLLIQMPILMAFYAIRGSSEI 121

5 Query: 164 SSATFLGLNLGQKSITLTVIIAILYFVQSWLSMQGVPDEQRQOMKTMMYLMEIMMVMSI 223
 +S TFL NLG + L +I ++Y Q ++SM G EQ++QMK + + PIM++F+S
 Sbjct: 122 ASHTFLWFNLGSPDMVLAIAGLVYLAQYFVSMIGYSPEQKKQMKIIGLMSFIMILFVSF 181

10 Query: 224 SLPASVALYWFIFGIFSIQQLVT--TYVLK-PKLRKVEEE 262
 + P+++ALYW +GG+F Q L+T Y+ K P+++ +EE
 Sbjct: 182 TAPSALALYWAVGGLFLAGQTLLTKKLYMNKHPEIKVMEQEE 223

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4789> which encodes the amino acid sequence <SEQ ID 4790>. Analysis of this protein sequence reveals the following:

15 Possible site: 31

>>> May be a lipoprotein

INTEGRAL	Likelihood = -9.55	Transmembrane	62 - 78 (54 - 82)
INTEGRAL	Likelihood = -2.81	Transmembrane	178 - 194 (177 - 195)
INTEGRAL	Likelihood = -0.90	Transmembrane	216 - 232 (215 - 232)

20 ----- Final Results -----

bacterial membrane	---	Certainty=0.4821(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

25

The protein has homology with the following sequences in the databases:

>GP:AAF03934 GB:AF139908 membrane protein homolog [*Listeria monocytogenes*]
 Identities = 89/218 (40%), Positives = 132/218 (59%), Gaps = 2/218 (0%)

30 Query: 43 KPMSYFIDYFANNAGLYGLAIIVTIIIVRTLILPLGLYQSWKASYQSEKMAFLKPVFEP 102
 +P + FI + A G YG+AIIII T+++R LI+PL L + KMA KP +
 Sbjct: 2 QPFTSFIMFVAKFVGGNYGIAIIITLLIRALIMPLNLRATAQMGMSKMAVAKPEIDE 61

35 Query: 103 INKRIKQANSQEEKMAAQTELMAAQRAHGINPLGGIGCLPLLIQMPFFSAMYFAAQYTKG 162
 I R+K+A S+EE+ Q E+MA + INP+ +GCLPLLIQMP A Y+A + +
 Sbjct: 62 IQARLKRATSKKEEQATIQKEMMAVYSKYNNINPMQ-MGCLPLLIQMPILMAFYAIRGSSE 120

40 Query: 163 VSTSTFMGIDLGSRSVLTAIIAALYFFQSWLSMMAVSEEQREQMKTMMYTMPIMMIFMS 222
 +++ TF+ +LGS +VL I +Y Q ++SM+ S EQ++QMK + PIM++F+S
 Sbjct: 121 IASHTFLWFNLGSPDMVLAIAGLVYLAQYFVSMIGYSPEQKKQMKIIGLMSFIMILFVS 180

Query: 223 FSLPAGVGLYWLVGGFSSIIQQLITTYLLKPRLHKQIK 260
 F+ P+ + LYW VGG F Q L+T L + H +IK
 45 Sbjct: 181 FTAPSALALYWAVGGLFLAGQTLLTKKLYMNK-HPEIK 217

An alignment of the GAS and GBS proteins is shown below.

Identities = 203/309 (65%), Positives = 254/309 (81%), Gaps = 2/309 (0%)

50 Query: 1 MKKTLKRILFSSLSLSMLLLLTGCVSVDKAGKPYGVINWTLGVPMANLITYFAHQGLGF 60
 +K TL RILFS L+LS+LL LTGCV D G P G+IW LG PM+ I YFA + GLG+
 Sbjct: 1 LKLTILNRILFSGIALSILLTLTGCVGRDAHGNPKGMIWEFLGKPMYSYFIDYFANNAGLY 60

55 Query: 61 GVAIIVTVIVRVVILPLGLYQSWKASYQAEKMAFYKPLFEPINERLRNAKTQEEKLAAQ 120
 G+AIIVT+IVR +ILPLGLYQSWKASYQ+EKMA+ KP+FEPIN+R++ A +QEEK+AAQ
 Sbjct: 61 GLAIIVTIIIVRTLILPLGLYQSWKASYQSEKMAFLKPVFEPINKRIKQANSQEEKMAAQ 120

60 Query: 121 TELMTAORENGLSMFGGIGCLPLLIQMPFFSAIFFAARYTPGVSSATFLGLNLGQKSLTL 180
 TELM AQR +G++ GGIGCLPLLIQMPFFSA++FAA+YT GVS++TF+G++LG +SL L
 Sbjct: 121 TELMAAQRHGINPLGGIGCLPLLIQMPFFSAMYFAAQYTKGVSTSTFMGIDLGSRSVL 180

Query: 181 TVIIAILYFVQSWLSMQGVPDEQRQOMKTMMYLMPIMMVMSISLPASVALYWFIFGIFS 240
 T IIA LYF QSWLSM V +EQR+QMKTMMY MPIMM+FMS SLPA V LYW +GG FS
 Sbjct: 181 TAIIAALYFFQSWLSMMAVSEEQREQMKTMMYTMPIMMIFMSFSLPAGVGLYWLVGGFSS 240

-1724-

```

732      762      792      822      852      894      918
FGGIGCLPLLIQMPFFSAIFFAARYTPGVSSATFLGLNLGQKSLTLTVIIAILYFVQSW---LSMQ--GVPDE--QRQQ
:||||:||||| :||| :|| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
-1AMGCLPMLIQSPIMIGLYAIRSTPEIASHSFLWFSLGQSDILMSLSAGIMYFVQAYIAQKLSAKYSAPVQNPAQQS
5      150      160      170      180      190      200      210

948      978      1008      1038      1068      1098      1128      1158
MKTMMYMLPIMMVFMSISLPASVALYWFIFGIFSIQQLVTTYVLKPKLRKVEEYTKNPPKAYKANNARKDVTNSTKA
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
10 AKLMVFIFPVMVTIFSLNVPALPLYWFTSGLFLTQVQNIQLMTHHKSCKTAALTESVK
      230      240      250      260      270

37.2/62.0% over 220aa
Listeria monocytogenes
15 GP|6117974| membrane protein homolog Insert characterized

ORF02470(430 - 1086 of 1530)
GP|6117974|gb|AAF03934.1|AF139908_4|AF139908(3 - 223 of 237) membrane protein homolog
{Listeria monocytogenes}
20 %Match = 14.6
%Identity = 37.1 %Similarity = 62.0
Matches = 82 Mismatches = 81 Conservative Sub.s = 55

285      315      345      375      405      435      465      495
K*NR*VY*KLEELKTLKKTTLKRILFSSLSLSMLLLLLTGCVSVDKAGKPYGVIWNTLGVPMANLITYFAHQHGLGFGVAII
| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
25 IQPFTSFIMFVAKFVGGNYGIAII
      10      20

525      555      585      615      645      675      705      735
IVTVIVRVVILPLGLYQSWKASYQAEKMAFYFKPLFEPINERLRNAKTQEEKLAAQTELMTAQRENGLSMFGGIGCLPLLI
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
30 ITTLLIRALIMPLNLRITAKAQMGMSKMAVAKPEIDEIQARLKRAATSKKEEQATIQEMMAVYSKYNNINP-MQMGCLPLLI
      40      50      60      70      80      90      100

765      795      825      855      885      915      945      975
QMPFFSAIFFAARYTPGVSSATFLGLNLGQKSLTLTVIIAILYFVQSWLSMQGVDPDEQRQOMKTMMYMLPIMMVFMSISL
||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
35 QMPILMAFYAIRGSSEIASHTFLWFLGSPDMVLAIITAGLVYLAQYFVSMIGYSPEQKKQMKIIGLMSPIMLFVSFTA
      120      130      140      150      160      170      180

1005      1035      1086      1116      1146      1176      1206
PASVALYWFIFGIFSIQQLVTT--YVLK-PKLRKVEEYTKNPPKAYKANNARKDVTNSTKATESNQAIITSKKTNRN
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
45 PSALALYWAVGGLFLAQTLTLTKLYMNKHPEIKVMEQEEKEFEQIVEEQKKEK
      200      210      220      230

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 1552

A DNA sequence (GBSx1644) was identified in *S.agalactiae* <SEQ ID 4791> which encodes the amino acid sequence <SEQ ID 4792>. This protein is predicted to be amino acid ABC transporter, permease protein. Analysis of this protein sequence reveals the following:

```

Possible site: 48
55 >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -9.98    Transmembrane    32 - 48 ( 23 - 53)
      INTEGRAL    Likelihood = -9.18    Transmembrane    195 - 211 ( 189 - 213)
      INTEGRAL    Likelihood = -8.70    Transmembrane    72 - 88 ( 62 - 93)

60 ----- Final Results -----
      bacterial membrane --- Certainty=0.4991(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-1725-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12131 GB:Z99105 similar to amino acid ABC transporter
(permease) [Bacillus subtilis]
Identities = 116/217 (53%), Positives = 168/217 (76%)

Query: 2 INWDAIFNLELAVKAPPSVIQGLPYTIGLSLVGFILGAIVGFFVALMKMSHFRLRLRYLAN 61
I W+ IFN +LA+++FP VI+G+ YT+ +S V G ++G F++L +MS LLR+ A
Sbjct: 5 IQWEYIFNTKLAIESFPYVIKIGIGYTLISFVSMFAGTVIGLFISLARMSKLALLRWPAK 64

Query: 62 IHISLMRGIPLMVLLFLIYFGLPFIGIQLDAVTASIVGFTMMSSAYISEIIRALLAVDH 121
++IS MRG+P++V+LF++YFG P+IGI+ AVTA+++GF++ S+AYI+EI R+A+ +V+
Sbjct: 65 LYISFMRGVPILVILFIFYFGFPYIGIEFSAVTAALIGFSLNSAAYIAEINRSAISSVEK 124

Query: 122 GQWEAARALGLKTPTIYRGIIIPQATRIALPSSLNVLLDMVKSSSLTAMITVPDIFNNAK 181
GQWEAA +LGL RGII+PQ+ RIALP L+NVLLD++K+SSL AMITVP++ +AK
Sbjct: 125 GQWEAASSLGLSYWQTMRGIIIPQSIRIALPPLANVLLDLIKASSLAAMITVPELLQHAQ 184

Query: 182 IVGGTYSYDYMAYILVALIYWVICTLYAIIQDWWEKR 218
I+GG DYMT YIL ALIYW IC++ A+ Q+ EK+
Sbjct: 185 IIGREFDYMTMYILTALIIYWAICSAAVFQNILEKK 221

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4793> which encodes the amino acid sequence <SEQ ID 4794>. Analysis of this protein sequence reveals the following:

Possible site: 23
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -6.79	Transmembrane	186 - 202 (184 - 205)
INTEGRAL	Likelihood = -5.84	Transmembrane	26 - 42 (21 - 43)
INTEGRAL	Likelihood = -4.78	Transmembrane	57 - 73 (56 - 84)
INTEGRAL	Likelihood = -1.59	Transmembrane	86 - 102 (86 - 103)

----- Final Results -----
bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB12131 GB:Z99105 similar to amino acid ABC transporter
(permease) [Bacillus subtilis]
Identities = 113/214 (52%), Positives = 157/214 (72%)

Query: 1 MINIPLMKDSLGFVLSGLPYTLGISLLSFFTGLFLGLGLALLGRSRQPLIHVLRAYISI 60
+ N L +S +V+ G+ YTL IS +S F G +GL ++L S+ L+ + + YIS
Sbjct: 10 IFNTKLAIESFPYVIKIGIGYTLISFVSMFAGTVIGLFISLARMSKLALLRWPAKLYISF 69

Query: 61 MRGVPMIVLVFLVLYFGLPYYGLELPALLCAYLGFSMVSAAYISEVFRSSIEAIDKGQWEA 120
MRGV++V+LF+LYFG PY G+E A+ A +GFS+ SAAYI+E+ RS+I +++KGQWEA
Sbjct: 70 MRGVPIVLVILFIFYFGFPYIGIEFSAVTAALIGFSLNSAAYIAEINRSAISSVEKGQWEA 129

Query: 121 AKALGLPYALMVKKIILPQAFRIAVPPLGNVIIDMVKSSSLAAMITVPDIFQNAKIIGR 180
A +LGL Y ++ IILPQ+ RIA+PPL NV++D++K+SSLAAMITVP++ Q+AKIIGR
Sbjct: 130 ASSLGLSYWQTMRGIIIPQSIRIALPPLANVLLDLIKASSLAAMITVPELLQHAKIIGR 189

Query: 181 EWDYMSMYILVAFIYWLI AFLERYQEFLNKLA 214
E+DYM+MYIL A IYW I + +Q LE K A
Sbjct: 190 EFDYMTMYIILTALIIYWAICSAAVFQNILEKKYA 223

An alignment of the GAS and GBS proteins is shown below.

Identities = 110/213 (51%), Positives = 156/213 (72%)

Query: 7 IFNLELAVKAPPSVIQGLPYTIGLSLVGFILGAIVGFFVALMKMSHFRLRLRYLANIHISL 66
+ N+ L + V+ GLPYT+G+SL+ F G +G +AL+ S L+ YL +IS+
Sbjct: 1 MINIPLMKDSLGFVLSGLPYTLGISLLSFFTGLFLGLGLALLGRSRQPLIHVLRAYISI 60

-1726-

Query: 67 MRGIPLMVLFLIYFGLPFFIGIQLDAVTASIVGFTMMSSAYISEIIRAALLAVDHGQWEA 126
 MRG+P++V+LF++YFGLP+ G++L A+ + +GF+M+S+AYISE+ R+++ A+D GQWEA
 Sbjct: 61 MRGVPMIVLVLFVLYFGLPYGLELPALLCAYLGFSMVSAAYISEVFRSSIEAIDKGQWEA 120

Query: 127 ARALGLKTPITYRGIIIPQATRIALPSLSNVLLDMVKSSSLTAMITVPDIFNNAKIVGGT 186
 A+ALGL + + II+PQA RIA+P L NV++DMVKSSSL AMITVPDIF NAKI+GG
 Sbjct: 121 AKALGLPYALMVKKIILPQAFRIAVPPLGNVIIDMVKSSSLAAMITVPDIFQNAKIIGGR 180

Query: 187 YSDYMTAYILVALIYVICTLYAIQDWEKRL 219
 DYM+ YILVA IYW+I L Q++ E +L
 Sbjct: 181 EWDYMSMYILVAFIYWLIAFLLERYQEFLENKL 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1553

A DNA sequence (GBSx1645) was identified in *S.galactiae* <SEQ ID 4795> which encodes the amino acid sequence <SEQ ID 4796>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> May be a lipoprotein

----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12132 GB:Z99105 similar to amino acid ABC transporter
 (binding protein) [Bacillus subtilis]
 Identities = 127/276 (46%), Positives = 183/276 (66%), Gaps = 12/276 (4%)

Query: 3 KTIILLGLVGLSAMTLAACS--NGQSSKETWTDNKKDGVLKVATPATLYPTSYYDDHK-- 58
 K ++ + LAACS N SK+T W+ IK G + VAT TLYPTS+D
 Sbjct: 8 KAVIFSFTMAFFLILAACSGKNEADSKDTGWEQIKDKGKIVVATSGTLYPTS+YHDTDSGS 67

Query: 59 -KLITGYEIDMMKAIAKKLKIKVFEVGVABSFSTVDSGKVDVAVNNFDTTPERLKKYNF 117
 KLITGYE++++ AK+L +KV+F E+G+ T+V+SG+VD A N+ D T +R +K+ F
 Sbjct: 68 DKLTGYEVEVVREAARLGLKVFKEMGIDGMLTAVNSGQVDAANDIDVTKDREEKFAF 127

Query: 118 SQFYKYSVGGMIVRADGSSKITAKDLSDWKGGKAGGGAGTQYMKIAKQQGAEPVIYDNVT 177
 S FYKYS G IVR D S I K L D KGKKA G A T YM++A++ GA+ VIYDN T
 Sbjct: 128 STFYKYSYGTAVRKDDLGI--KTLKDLKGKKAAGAATTVMVARKYGAKEVIYDNAT 185

Query: 178 NDVYLRLDVSTGRDIFIPNDYYTQVIADVVKYVTKQYVDIKVKM-GDVKNPTEQGIVMSKKD 236
 N+ YL+DV+ GRD I NDYY Q +A+ +PD+ + + D+KY P +Q +VM K +
 Sbjct: 186 NEQYLKDVANGRTDVILNDYYLQTLAL---AAPDLNITIHPDIKMPNKQALVMKKS 241

Query: 237 KSLKTKIDAAIKDMKDGSLKKISEKYYAGQDLTKE 272
 +L+ K++ A+K+M KDGLS K+S++++ D++K+
 Sbjct: 242 AALQKKMNEALKEMSKDGLTKLSKQFFNKADVSKK 277

There is also homology to SEQ ID 1190.

SEQ ID 4796 (GBS183) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 2; MW 33kDa).

GBS183-His was purified as shown in Figure 199, lane 7.

-1727-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1554

A DNA sequence (GBSx1646) was identified in *S.galactiae* <SEQ ID 4797> which encodes the amino acid sequence <SEQ ID 4798>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1514(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAF09821 GB:AE001885 6-aminoheptanoate-cyclic-dimer hydrolase
[Deinococcus radiodurans]
Identities = 178/488 (36%), Positives = 265/488 (53%), Gaps = 17/488 (3%)

20 Query: 5 DATAMVQAIKQHKISSQELVEQAIYKIEEQNVSVNAVVSQYNEARQAAYANESNA--- 61
DA + Q ++ ++S++++ AI++ + NV++NAV V Y++ A+ + + A
Sbjct: 54 DALDLAQLFRRGELSAEDMCTAAIHRAQVNVVALNAV VYPLYDQGLAQARATDAARARGE 113

25 Query: 62 ----PFAGVPILLKDLGQNKQQLSTSGSQLFKHYHAKQTDYLVQSFEKLGFIILGRNT 117
PFAGVP L+KD G G T G++ ++ + D LV+ ++ G + LG+TNT
Sbjct: 114 QATGPFAGVPFLVKDFGSRLAGVPHTGGTRAYRQIPEWDDLVRRWQAAGLLPLGKTNT 173

30 Query: 118 PEFQFKNISDQQLHGNVNLFPDHSRNAGGSSGGAAAAVSSGMVPIAGSDGGGSIRIPAS 177
PEF +++ +LHG P+D R GGSSGG+A+AV++G+VP+AGA DGGGSIRIPAS
Sbjct: 174 PEFALMGVTEPELHGPTNPWDLGRTPGGSSGGSASAVAAGIVPLAGAGDGGGSIRIPAS 233

35 Query: 178 FNGLIGLKPSRGRIPVGPSSYRGWQASSHFALTKSVRDTKRLLYYLQSYQVES----PF 233
GL GLKPSRGR+P G WQGA+ LT+SVRD+ LL Q + P
Sbjct: 234 CCGLFGLKPSRGRVPCGDGVGEPWQGAAVEHVLTRSVRDSAALLDLEQGPDAAGALFLPS 293

40 Query: 234 PLKKLSKESLFEFSVSKPLKIAVLMDSPKTKVSSEAKAIKEADFLSQGNHLELVEQ 293
P + S+E E L+I PL V E AA++ AA L G+ +E V
Sbjct: 294 PERPYSEEVGRE---PGRLRIGFSTAHLGRSVHPECVAAVQGAARLLES LGHEVEEVAL 350

45 Query: 294 PLDGIHSMKTYCMMNSVETAAMFDDIEKSLGRSMEFSDMELMTWAMYQSGQRLAKDYK 353
P DG + + M+ ET A + +LGR SD+E +TW + Q G+ A D++
Sbjct: 351 PWDGPALAQAFLLMYFGETGASLAALRDTLGRPARASDVEAVTWLLGQLGRSYSAADFAA 410

50 Query: 354 LLDSWDQFAATMARFHENYDLILTAATNQAPFHHGQFD---LDETLQKQLRHMGEFSVSE 410
SW+ A M RFH+NYDL+LT P G+ + L + + M +
Sbjct: 411 ARASWNVHARAMGRFHQNYDLLLLTPVLATPPLQIGELQPRGVQAALLRAAQMDVSGLLR 470

55 Query: 411 QQDLIWKMFEDSMAWTPTFTHQPNLTGQPSLAIPHTLTKEGLPLGVQLTAAKGREDLLLA 470
+ + + D + P+T NLTGQP++++P H T +GLP+GVQ A RED+LL +
Sbjct: 471 RSGQVDALATDILEKMPYQLANLTGQPAMSVPLHWTADGLPVGQVFVAPLAREDVLLRL 530

Query: 471 AELFEKEK 478
A E+ +
Sbjct: 531 AGQLEQAR 538

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4047> which encodes the amino acid sequence <SEQ ID 4048>. Analysis of this protein sequence reveals the following:

Possible site: 15
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

60

-1728-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 277/484 (57%), Positives = 348/484 (71%), Gaps = 2/484 (0%)

Query: 1 MVFKDATAMVQAIKQHKISSQELVEQAIYKIEEQNVSVNAVVSQYNEARQAAKYANESN 60
 M ++DATAM A++ + + ELV QAIYK ++ N ++NA+ S+++ A + AK + S

10 Sbjct: 1 MTYQDATAMAIQVQTGQTTPLELVTAIYKAKKLNPTLNATTSERFEAALEEAKQDFSG 60

Query: 61 APFAGVPILKDLGQNKQKQLSTSGSGLFKHYHAKQTDYLVQSFEKLGFIILGRNTPTPEF 120
 PFAGVP+ LKDLGQ KG STSGS+LFK Y A +TD V+ E LGFIILGR+NTPEF

15 Sbjct: 61 LPFAGVPLFLKDLGQELKGHSSTSGSRLFKYQATKTDLFVKRLEALGFIILGRSNTPTPEF 120

Query: 121 GFKNISDGQLHGNVNLFPDHSRNAGSSSGGAAAVSSGMVPIAGASDGGGSIRIPASFNG 180
 GFKNISD LHG VNLP D++RNAGSSSGGAAA VSSG+ +A ASDGGGSIRIPASFNG

Sbjct: 121 GFKNISDSSLHGPVNLPDNRNAGSSSGGAAALVSSGISALATASDGGGSIRIPASFNG 180

20 Query: 181 LIGLKPSRGRIPVGPSSYRGWQGASSHFALTKSVRDTRKLLYYLQSYQVESPFPLKKLSK 240
 LIGLKPSRGR+PVGP SYR WQGAS HFALTKSVRD+ LLYYLQ Q+ESPFPL L+K

Sbjct: 181 LIGLKPSRGRMPVGPSSYRSWQGASVHFALTKSVRDTRNLLYYLQMEQMESPFPLATLTK 240

Query: 241 ESLFEFSVSKPLKIAVLMDSPKTKVSSEAKAAIKEAADFLSQKGNHL-ELVEQPLDGIH 299
 +S+++ S+ +PL IA + VS + A+++A +L ++G+ L EL E P++

25 Sbjct: 241 DSIYQ-SLQRPLTIAFYQRLSDGSPVSLDTAKALRQAVTWLREQHQLVELEEFVNMTE 299

Query: 300 SMKTYCMMNSVETAAMFDDIEKSLGRSMESDMELMTWAMYQSGQRVLAKDYSKLLDSWD 359
 ++ Y +MNSVETAAMF DIE + GR M DME MTWA+YQSG+ + A YS++L WD

30 Sbjct: 300 VIRHYIMNSVETAAMFADIEDTFGRPMTKDDMETMTWAIYQSGKDIPAWRYSQVLQKWD 359

Query: 360 QFAATMARFHENYDLILTAATNQAPPFHGQFDLDETLOQLRHMGEFSVSEQODLIWKMF 419
 ++ATMA FHE YDL+LT TN PAP HG+ D L L FS EQ +L+ MF

35 Sbjct: 360 TYSATMASFHETFDLLTFTTNTPAKHHGELVPDSKLMANLAQAEIFSSEEQFNLVETMF 419

Query: 420 EDSMAWTPFTHQPNLTGQPSLAIPHTLTKEGLPLGVQLTAAKGREDLLLAVAELEFEKEKQ 479
 S+A P+T PNLTGQP++++PT+ TKEGL +G+QL AAKGREDLLL +AE FE

Sbjct: 420 GKSLAINPYTALPNLTGQPAISLPTYETKEGLSMGIQLIAAKGREDLLLGIAEQFEAAGL 479

40 Query: 480 FKGP 483
 K P

Sbjct: 480 LKIP 483

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 45 vaccines or diagnostics.

Example 1555

A DNA sequence (GBSx1647) was identified in *S. agalactiae* <SEQ ID 4799> which encodes the amino acid sequence <SEQ ID 4800>. This protein is predicted to be transcription elongation factor (greA). Analysis of this protein sequence reveals the following:

50 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.5003 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

60 >GP:CAB14674 GB:Z99117 transcription elongation factor [Bacillus subtilis]
 Identities = 86/154 (55%), Positives = 114/154 (73%), Gaps = 1/154 (0%)

-1729-

Query: 3 EKTYPMTQVEKDQLEKELEELKLVRRPEVVERIKIARSYGDLSSENSEYDAAKDEQAFVEG 62
 EK +PMT K +LE+ELE LK V+R EVVERIKIARS+GDLSSENSEYD+AK+EQAFVEG
 Sbjct: 4 EKVPFMTAEGKQKLEQELEYLKTVKRKEVVERIKIARSFGDLSSENSEYDSAKEEQAFVEG 63

Query: 63 QIQILETKIRYAEIIDSDAVAKDEVAIGKTVLVQEVGTNDKDTYHIVGAAGADIFSGKIS 122
 ++ LE IR A+II+ D + V +GKTV E+ D+++Y IVG+A AD F GKIS
 Sbjct: 64 RVTTLNMRNAKIIEDDG-GSNVVLGKTVTFVELPDGDEESYTIVGSAEADPFEGKIS 122

Query: 123 NESPIAHALIGKKTGDLATIESPAGSYQVEIISV 156
 N+SPIA +L+GKK + T+++P G V+I+ +
 Sbjct: 123 NDSPIAKSLLGKKVDEEVTVQTPGGEMLVKIVKI 156

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4801> which encodes the amino acid sequence <SEQ ID 4802>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4434(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 145/160 (90%), Positives = 149/160 (92%)

Query: 1 MAEKTYPMTQVEKDQLEKELEELKLVRRPEVVERIKIARSYGDLSSENSEYDAAKDEQAFV 60
 MAEKTYPMT EK+QLEKELEELKLVRRPE+VERIKIARSYGDLSSENSEYDAAKDEQAFV
 Sbjct: 17 MAEKTYPMTLTEKEQLEKELEELKLVRRPEIVERIKIARSYGDLSSENSEYDAAKDEQAFV 76

Query: 61 EGQIQILETKIRYAEIIDSDAVAKDEVAIGKTVLVQEVGTNDKDTYHIVGAAGADIFSGK 120
 EGQI LETKIRYAEIIDSDAVAKDEVAIGKTV+VQEVGT DKDTYHIVGAAGADIFSGK
 Sbjct: 77 EGQISTLETKIRYAEIIDSDAVAKDEVAIGKTVIVQEVGTTDKDTYHIVGAAGADIFSGK 136

Query: 121 ISNESPIAHALIGKKTGDLATIESPAGSYQVEIISVEKTN 160
 ISNESPIA ALIGKKTGD IESPA +Y VEIISVEKTN
 Sbjct: 137 ISNESPIAALIGKKTGDKVRIESPAATYDVEIISVEKTN 176

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1556

A DNA sequence (GBSx1648) was identified in *S.agalactiae* <SEQ ID 4803> which encodes the amino acid sequence <SEQ ID 4804>. This protein is predicted to be aminodeoxychorismate lyase-like protein. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -13.64 Transmembrane 238 - 254 (230 - 260)

----- Final Results -----
 bacterial membrane --- Certainty=0.6456(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF77615 GB:AF151720 aminodeoxychorismate lyase-like protein
 [Streptococcus thermophilus]
 Identities = 135/210 (64%), Positives = 171/210 (81%)

-1730-

Query: 373 KTTSTPYKADDFLKLVDQETFIKMKVAKYPNLLGSLPDKSKAIYQLEGYLFPPATYNYKYD 432
 K +ST K DFLKL++D+ FI KM AKYP LL +LP+ + A Y LEGYLFPPATYN + D
 Sbjct: 5 KHSSTGLKEKDFLKLMDDAFITKMKAKYPTLLANLPNSTDAKYVLEGYLFPPATYNIHDD 64

5 Query: 433 TTLEGLVEDMISTMTKMAPYNTIKAKNMSVNDVLTSSLVEKEGSTDEDRRKIASVVFY 492
 TT+E L E+M+ TM+T ++PYY TI + N +VN++LTL+SLVEKEG+TD+DR+ IASVVFY
 Sbjct: 65 TTVESLAEEMLFMTDTHLSPYYATILSSNHNVEILTLASLVEKEGATDDDRKNIASVVFY 124

10 Query: 493 NRLSAGQALQSNIAILYAMGKLGDKTSLAEDAQINTSIKSPYNIYTNTGLMPGPVDSPSI 552
 NRL++ ALQSNIA+LY +GKLG +T+L EDA I+T+I SPYN Y + GLMPGPVDSPS+
 Sbjct: 125 NRLNSDMALQSNIAVLVYVLGKLGQETTLKEDATIDTNIDSPYNDYVHKGLMPGPVDSPSL 184

Query: 553 SAIBATIKPASTDYLYFVADVKTGNVYVYAK 582
 SAIEA I P+ST Y+YFVADV TGNVY+A+
 15 Sbjct: 185 SAIEAVINPSSTKYMYFVADVSTGNVYFAE 214

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4805> which encodes the amino acid sequence <SEQ ID 4806>. Analysis of this protein sequence reveals the following:

Possible site: 59

20 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -7.91 Transmembrane 161 - 177 (155 - 183)

----- Final Results -----

25 bacterial membrane --- Certainty=0.4163 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

30 >GP:AAF77615 GB:AF151720 aminodeoxychorismate lyase-like protein
 [Streptococcus thermophilus]
 Identities = 135/212 (63%), Positives = 161/212 (75%)

35 Query: 295 KTKKAKTFNEKDFLDLVTDEAFIQDMVKRYPKLLATIPTKEKAIYRLEGYLFPPATYNY 354
 K K + T EKDFL L+ D+AFI M +YP LLA +P A Y LEGYLFPPATYN +
 Sbjct: 3 KGHSSSTGLKEKDFLKLMDDAFITKMKAKYPTLLANLPNSTDAKYVLEGYLFPPATYNIH 62

Query: 355 KETTMRELVEDMLAAMDATLVPPYDKIAASGKIVNEVLTSLASLVEKEGSTDDDRRQIASV 414
 +TT+ L E+ML MD L PYY I +S VNE+LTLASLVEKEG+TDDDR+ IASV
 40 Sbjct: 63 DDTTVESLAEEMLFMTDTHLSPYYATILSSNHNVEILTLASLVEKEGATDDDRKNIASV 122

Query: 415 FYNRLNSGMALQSNIAILYAMGKLGKTTLAEDATIDTTINSPYNIYTNTGLMPGPVASS 474
 FYNRLNS MALQSNIA+LY +GKLG++TTL EDATIDT I+SPYN Y + GLMPGPV S
 45 Sbjct: 123 FYNRLNSDMALQSNIAVLVYVLGKLGQETTLKEDATIDTNIDSPYNDYVHKGLMPGPVDS 182

Query: 475 GVSALATLNPASTDYLYFVANVHTGEVYVYAK 506
 +SAIEA +NP+ST Y+YFVA+V TG VY+A+
 Sbjct: 183 SLSAIEAVINPSSTKYMYFVADVSTGNVYFAE 214

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 310/603 (51%), Positives = 403/603 (66%), Gaps = 86/603 (14%)

Query: 1 MTEFNDDQHSNHDQKSFKEQILAELEENRLRLREEELYQKEQEAKEAARRTAQLMADY 60
 +T+F D + Q+SFKEQILAELE+AN++RK +EEL+
 55 Sbjct: 3 LTDFKDKDQDQ-QRSFKEQILAELEKANQIRKEKEEELF----- 41

Query: 61 EAQRLKDEREARAKALETQRLBQEKARIEAKLLAEAREEERRQAEQALASQEEQVIN 120
 ++ LE +E AR A+L AE R++ A Q+E + +
 60 Sbjct: 42 -----QKELEAKAARRTAQLYAEYKRQD-----AFQKESIAH 74

Query: 121 QGMEPSRELDGSGKSSEFRITTENVPDIDLKADKTDVATAVPNQETEEIFLVRATDIPTG 180
 +T ++ +A K V T+ + T + +E
 Sbjct: 75 NN-----KTAKH-----PQAIGAVMTSEALKPT-----LLSEK 103

-1731-

Query: 181 ENVKLGEISELEPVAKEPIRVEDLSKEEGIALSAKNKHNRER---RQKADNVAKRIAR 237
 EN L ++ A E +++ + +E + L+ + H+ R + RQ+ + AK+I+
 Sbjct: 104 ENSSLKTTNKRVRVQANE---LQETASKESQVPLTIEKGHSVRRKLSKRQQTERAAKKIST 160

5 Query: 238 ILISIIILVLLLTAFVGYRFVDSAIKPVDSNSNKFVQVEIPIGSGNKLIGQILEKAGVIK 297
 +LIS II+ LL G +V SA+ PVD NS+ FVQVEIP GSGNKLIGQIL+K G+IK
 Sbjct: 161 VLISSIIITLLAVTLAGAGYVYSALNPFVDKNSDAFVQVEIPSGSGNKLIGQILQKKGLIK 220

10 Query: 298 SATVFNYYSKFKNYSNFQSGYYNLKKSMTILDQIAAELEKGGTAEPTKPALGKILITEGYT 357
 ++TVF++Y+KFKN++NFQSGYYNL+KSM+L++IA+ L++GGTAEPTKP+LGKILI EGYT
 Sbjct: 221 NSTVFSFYTKFKNFTNFQSGYYNLQKMSLEEIASALQEGGTAEPTKPSLGKILIEGYT 280

15 Query: 358 IKQIAKAIESN-KIDTKTTSTPYKADDFLKLVDQETFIKKMVAKYPNLLGSLPDKSKAIY 416
 IKQIAKA+E N K TK TP+ DFL LV DE FI+ MV +YP LL ++P K KAIY
 Sbjct: 281 IKQIAKAVEHNSKGKTKKAKTPFNEKDFLDLVTDEAFIQDMVKRYPKLLATIPTKEKAIY 340

20 Query: 417 QLEGYLFPATYNYKDTTLEGLVEDMISTMTKMAPYYNTIKAKNMSVNDVLTSSLVEK 476
 +LEGYLFPATYNYK+TT+ LVEDM++ M+ + PYY+ I A +VN+VLT+SLVEK
 Sbjct: 341 RLEGYLFPATYNYKETTRELVEDMLAAMDATLVPPYDKIAASGKTVEVLTSLASVEK 400

25 Query: 477 EGSTDEDRRKIASVFYNRLSAGQALQSNIAILYAMGKLGDKTSLAEDAQINTSIKSPYNI 536
 EGSTD+DRR+IASVFYNRL++G ALQSNIAILYAMGKLG+KT+LAEDA I+T+I SPYNI
 Sbjct: 401 EGSTDDRRQIASVFYNRLNSGMALQSNIAILYAMGKLGKKTTLAEDATIDTTINSYNI 460

30 Query: 537 YNTTGLMPGPVDSPSISAEATIKPASTDYLYFVADVKTGNVYAKDFETHKANVEKYIN 596
 YNTTGLMPGPV S +SAIEAT+ PASTDYLYFVA+V TG VYAK FE H ANVEKY+N
 Sbjct: 461 YNTTGLMPGPVASSGVSAIEATLNPASTDYLYFVANVHTGEVYAKTFEEHSANVEKYVN 520

Query: 597 SQI 599
 SQI
 Sbjct: 521 SQI 523

A related GBS gene <SEQ ID 8843> and protein <SEQ ID 8844> were also identified. Analysis of this protein sequence reveals the following:

35 Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: -17.88
 GvH: Signal Score (-7.5): -3.51
 Possible site: 58
 >>> Seems to have no N-terminal signal sequence

40 ALOM program count: 1 value: -13.64 threshold: 0.0
 INTEGRAL Likelihood = -13.64 Transmembrane 238 - 254 (230 - 260)
 PERIPHERAL Likelihood = 5.78 285
 modified ALOM score: 3.23

45 *** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.6456(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF00931(1417 - 2046 of 2400)
 GP|8574530|gb|AAF77615.1|AF151720.1|AF151720(5 - 214 of 214) aminodeoxychorismate lyase-
 55 like protein {Streptococcus thermophilus}
 %Match = 17.5
 %Identity = 64.3 %Similarity = 81.4
 Matches = 135 Mismatches = 39 Conservative Sub.s = 36

60 1236 1266 1296 1326 1356 1386 1416 1446
 NYYSKFKNYSNFQSGYYNLKKSMTILDQIAAELEKGGTAEPTKPALGKILITEGYTIKQIAKAIESNKIDTKTTSTPYKAD
 |.:|||
 AKKGKHSSTGLKEK
 10

65

-1733-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 87/159 (54%), Positives = 117/159 (72%), Gaps = 1/159 (0%)

10 Query: 29 MIIRNGCLEDLQQVISIEQINFSEAEAAASKKAMQERLTIMTDTFLVAEINGRLAGYIEGP 88
 M+IR DL+ + +IE NFS EA ++ ++E + ++ DTFLVA I+ + GYIEGP
 Sbjct: 1 MLIRQVQGSDELVIATIESDNFSPQEQATTRAFLVEEHIRLIPDTFLVALIDQEIYGYIEGP 60

15 Query: 89 VIKGRYLTDDLPHKVSEFPVRVGGFIGITSLSIHPDFKGGIGTALLAAMKDLVVSQERD 148
 V+ L D LFH V++ P + GG+I ITSLSI F+ QG+GTALLAA+KDLVV+Q+R
 Sbjct: 61 VVTTFPIEDSLPHGVTKNP-KTGGYIAITSLSIKHFQQGGVGTALLAALKDLVVAQQRT 119

Query: 149 GISLTCHDDLISFYEMNGFKDEGESDSKHGGSLWYNMIW 187
 G+ LTCHD LIS+YEMNGF ++G S+S+HGG+LWY MIW
 20 Sbjct: 120 GLILTCHDYLLISYEMNGFINQGISQHGGLWYQMIW 158

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1558

25 A DNA sequence (GBSx1650) was identified in *S.agalactiae* <SEQ ID 4811> which encodes the amino acid sequence <SEQ ID 4812>. This protein is predicted to be udp-n-acetylmuramate--alanine ligase (murC/ddIA). Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have a cleavable N-term signal seq.

30 INTEGRAL Likelihood = -2.60 Transmembrane 272 - 288 (270 - 288)

----- Final Results -----

35 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00294 GB:AF008220 putative UDP-N-acetylmuramate-alanine
 ligase [Bacillus subtilis]

40 Identities = 238/432 (55%), Positives = 315/432 (72%), Gaps = 3/432 (0%)

Query: 5 YHFIGIKSGMSALALMLHQMGNVQGSDDVDKYYFTQGLEQAGVTILPFSFNNISEDLE 64
 YHF+GIKG+GMS LA +LH G+ VQGSDD++K+ FTQ LE+ +TILPFS NI +
 Sbjct: 4 YHFVGIKGTGMSPLAQILHDNGYTVQGSDDIEKFIFTQTALEKRNTILPFSFSAENIKPGMT 63

45 Query: 65 IIAGNAFRPDNNEELAYVIEKGQFKRYHEFLGDFMRQFTSLGVAGAHGKTSTTGLLAHV 124
 +IAGNAF PD + E+ + +G RYH+FLGD+M++FTS+ V GAHGKTSTTGLLAHV
 Sbjct: 64 VIAGNAF-PDTHPETEKAMSEGIPVIRYHKFLGDYMKKFTSVAVTGAHGKTSTTGLLAHV 122

50 Query: 125 LKNITDTSFLIGDGTGRGSANANYFVFEADYERHFMFYHPEYSIITNIDFDHPDYFTGL 184
 ++N TSFLIGDGTG+G+ N+ YFVFEA EY RHF+ Y P+Y+I+TNIDFDHPDYF+ +
 Sbjct: 123 IQNAKPTSFLIGDGTGGNENSEYFVFEACEYRRHFLSYQPDYAIMTNIDFDHPDYFSSI 182

55 Query: 185 EDVFNAFNDAKQVQKGLFIYGEDPKLHEITSEAPIYYYGFEDSNDFIKADITRTVNGSD 244
 +DVF+AF + A QV KG+ G+D L +I + P+ YYG + NDF A++I ++ G+
 Sbjct: 183 DDVFDAFQEMALQVNKGIIACGDDEHLPKIHANVPVYVYGTGEENDFQARNIVKSTEGTT 242

Query: 245 FKVFYNNQEEIGQFHVYPAYGKHNLNATAVIANLYINGIDMALVAEHLKTFSGVKRRFTEK 304
 F VF F++PAYG HN+LN+ AVIA + ID +++ LK+F GVKRRF EK
 60 Sbjct: 243 FDFVFRNTFYDTFYIPAYGHHNVLSLAVIALCHYEEIDSSIIKHALKSFGGVKRRFNEK 302

-1734-

Query: 305 IIDDTVIIDDFAHHPTEIATLDAARQKYPskeivaifQPHtFTRTIALLLDEFahalsQA 364
 + D V+IDD+AHHPTEI T++AARQKYP +EIVA+QPHtFTRT LDEFA +LS A
 SbJct: 303 QLGDQVLIDDYAHHPTEIKVTIEAARQKYPDREIVAVFQPHtFTRTQQFLDEFAESLSGA 362

5

Query: 365 DSVYLAQIYGSAREVDNGEVKVEDLAAKIVKHSdLVTVENVSPLLNHDNAVYVFMGAGDI 424
 D VYL I+GSARE + G++ + DL KI ++ L+ ++ S L HD AV +FMGAGDI
 SbJct: 363 DCVYLCDFGSARE-NAGKLtIGDLQgKI-HNAKLIEEDDTSVLKAHDKAVLIFMGAGDI 420

10

Query: 425 QLYERSFEELLA 436
 Q Y R++E ++A
 SbJct: 421 QKYMRAyENVMA 432

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4813> which encodes the amino acid
 15 sequence <SEQ ID 4814>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -4.57 Transmembrane 271 - 287 (269 - 288)

20

----- Final Results -----
 bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:AAC00294 GB:AF008220 putative UDP-N-acetylmuramate-alanine
 ligase [Bacillus subtilis]
 Identities = 236/431 (54%), Positives = 310/431 (71%), Gaps = 2/431 (0%)

30

Query: 5 YHFIGIKGSGMSALALMLHQMGHKVQGSdVEKYyFTQRGLEQAGITILPFSEDNITPDME 64
 YHF+GIKG+GMS LA +LH G+ VQGSd+EK+ FTQ LE+ ITILPFS +NI P M
 SbJct: 4 YHFVGIKGtGMSPLAQILHDNGYTVQGSdIEKFI FTQTALeKRNIITILPFSaENIKPGMT 63

35

Query: 65 LIVGNAFRENKKEVAYALRHQIPFKRYHDFLGDFMKSFISFAVAGAGKtSTTGLLSHVL 124
 +I GNAF + + E+ A+ IP RYH FLGD+MK F S AV GAHGKtSTTGLL+HV+
 SbJct: 64 VIAGNAFPDTHPEIKAMSEGIPVIRYHKFLGDYMKKFTSVAVTGAGKtSTTGLLAHVI 123

40

Query: 125 KNITDTSYLIGDGTGRGSANAQYVFESDEYERHFMPYHPEYSIITNIDFDHPDYFTGIA 184
 +N TS+LIGDGTG+G+ N++YVFVE+ EY RHF+ Y P+Y+I+TNIDFDHPDYF+ I
 SbJct: 124 QNAKPTSFILIGDGTGQGNENSEYFVFEACEYRRHFLSYQPDYAIMTNIDFDHPDYFSSID 183

45

Query: 185 DVRNAFN DYAKQVKKALFVYGEDDELKKIEAPAPIYyyGFEEGNDFIAYDITRTTNGSDF 244
 DV +AF + A QV K + G+D+ L KI A P+ YYG E NDF A +I ++T G+ F
 SbJct: 184 DVFDAFQEMALQVNGIITACGDDEHLPKIHANVPVYVYGTGEENDFQARNIVKSTEGTTF 243

50

Query: 245 KVKHQGEVIGQFHVFPAYGKHNLNATAVIANL FVAGIDMALVADHLKTFSGVKRRFTEKI 304
 V + F++PAYG HN+LN+ AVIA ID +++ LK+F GVKRRF EK
 SbJct: 244 DVFVRNTFYDTFYIPAYGHHNVLNSLAVIALCHYEIDSSIIKHALKSFGGVKRRFNEKQ 303

55

Query: 305 INDITIIIDDFAHHPTEIVATIDAARQKYPskeivaifQPHtFTRTIALLEDFAcALNEAD 364
 + D ++IDD+AHHPTEI TI+AARQKYP +EIVA+QPHtFTRT L++FA +L+ AD
 SbJct: 304 LGDQVLIDDYAHHPTEIKVTIEAARQKYPDREIVAVFQPHtFTRTQQFLDEFAESLSGAD 363

60

Query: 365 SVYLAQIYGSAREVDKGEVKVEDLAAKIIKPSQVVTVENVSPLLDHDNAVYVFMGAGDIQ 424
 VYL I+GSARE + G++ + DL K I ++++ ++ S L HD AV +FMGAGDIQ
 SbJct: 364 CVYLCDFGSARE-NAGKLtIGDLQgK-IHNAKLIEEDDTSVLKAHDKAVLIFMGAGDIQ 421

Query: 425 LYHSFEELLA 435
 Y ++E ++A
 SbJct: 422 KYMRAyENVMA 432

An alignment of the GAS and GBS proteins is shown below.

Identities = 369/443 (83%), Positives = 406/443 (91%), Gaps = 1/443 (0%)

-1735-

Query: 1 MSKTYHFIGIKGSGMSALALMLHQMGNVQGSVDKYYFTQRGLEQAGVTILPFSPNNIS 60
 MSKTYHFIGIKGSGMSALALMLHQMGNH VQGS DV+KYYFTQRGLEQAG+TILPFS +NI+
 Sbjct: 1 MSKTYHFIGIKGSGMSALALMLHQMGNH VQGS DVKEKYYFTQRGLEQAGITILPFSEDNIT 60

5 Query: 61 EDLEIIAGNAFRPDNNEELAYVIEKGYQFKRYHEFLGDFMRQFTSLGVAGAHGKTSTTGL 120
 D+E+I GNAFR +NN+E+AY + F KRYH+FLGDFM+ F S VAGAHGKTSTTGL
 Sbjct: 61 PDMEILVGNAPR-ENNKEVAYALRHQIPFKRYHDFLGDPMKSFISFAVAGAHGKTSTTGL 119

10 Query: 121 LAHVLKNITDTSFLIGDGTGRGSANANYFVFEADYERHFMPIHPEYSIITNIDFDHPDY 180
 L+HVLKNITDTS+LIGDGTGRGSANA YFVFE+DEYERHFMPIHPEYSIITNIDFDHPDY
 Sbjct: 120 LSHVLKNITDTSYLLIGDGTGRGSANAQYFVFESEYERHFMPIHPEYSIITNIDFDHPDY 179

15 Query: 181 FTGLEDFVNAFNNDYAKQVQKGLFTYGEDPKLHEITSEAPIYYYGFEDSNDFIADITRTV 240
 FTG+ DV NAFNDYAKQV+K LF+YGED +L +I + APIYYYGFE+ NDFIA DITRT
 Sbjct: 180 FTGIADVRNAFNNDYAKQVKKALFVYGEDDELKKEAPAPIYYYGFEEGNDFIAYDITRTT 239

20 Query: 241 NGSDFKVFNQEEIGQFHPAYGKHNLNATAVIANLYIMGIDMALVAEHLKTFSGVKRR 300
 NGSDFKV + E IGQFHPAYGKHNLNATAVIANL++ GIDMALVA+HLKTFSGVKRR
 Sbjct: 240 NGSDFKVKHQGEVIGQFHPAYGKHNLNATAVIANLFVAGIDMALVADHLKTFSGVKRR 299

25 Query: 301 FTEKIIDDIVIIDDFAHHPTEIITLDAARQKYPskeivaIFQPHFTFRTIALLDEFABA 360
 FTEKII+DT+IIDDFAHHPTEI+AT+DAARQKYPskeivaIFQPHFTFRTIALL++FA A
 Sbjct: 300 FTEKIINDTIIIDDFAHHPTEIVATIDAAARQKYPskeivaIFQPHFTFRTIALLEDFAA 359

30 Query: 421 AGDIQLYERSFEELLANLTKNIQ 443
 AGDIQLYE SFEELLANLTKN Q
 Sbjct: 420 AGDIQLYEHSFEELLANLTKNQ 442

SEQ ID 4812 (GBS157) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 24 (lane 11; MW 49kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 8; MW 74kDa), Figure 33 (lane 8; MW 74kDa) and Figure 37 (lane 3; MW 74kDa).

The GBS157-GST fusion product was purified (Figure 112A; see also Figure 200, lane 3) and used to immunise mice (lane 1+2 product; 19.5µg/mouse). The resulting antiserum was used for Western blot (Figure 112B), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

SEQ ID 4812 (GBS157) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 183 (lane 11-13; MW 74kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 1559

A DNA sequence (GBSx1651) was identified in *S.galactiae* <SEQ ID 4815> which encodes the amino acid sequence <SEQ ID 4816>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1980(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55

-1736-

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4817> which encodes the amino acid sequence <SEQ ID 4818>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2731(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 80/201 (39%), Positives = 126/201 (61%), Gaps = 9/201 (4%)

Query: 7 RFPLIADDEPVMSPVLKMNLYDNEDLINNIRDFYQEKTYQSMVKSNEYEHSEISHPKVIEN 66
+FPL+AD + P +M LY+NEDLI NIR +YQ+K Y + ++ EE +
Sbjct: 5 QFPLVADGIATSDPAKQMALYENEDLITNIRGVYQDKEYDDIARN-----EFTAKATSRQ 60

Query: 67 DPVPPQ--SFVKKATELSKSRQEAQRSVREKRQAYYAKQEFKAPSKEAFQQQLKATVPKK 124
P + S +K + ++RQ+AK+ ++EKQAY AK+ P + + +QQ + P +
Sbjct: 61 TPSSKRFCSNDEKHHYVKEARQAKQDLKEKRQAYLAKEMAYVPKQVSKKQPADSSPSQ 120

Query: 125 QTQRKVTLSHLSDRLQQESYILAEIPIIFQEPDNTPNP-KTKKNNFDLKRQVYNKQD 183
+ + TE+S + +L Q++YILAE+P ++EP N P TTKNN+DFLK SQ+YN ++
Sbjct: 121 K--QATTEMSRFTKKLHQDNYILAEPLKEYKEPKNLEQQGTTTKNNYDFLKSSQIYNNKE 178

Query: 184 NQFHKEKAQELNLTRFKDI 204
+ +E+ AQELNL+RF+D+

Sbjct: 179 MRQREKTIAQELNLSRFEDL 199

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1560

A DNA sequence (GBSx1652) was identified in *S.agalactiae* <SEQ ID 4819> which encodes the amino acid sequence <SEQ ID 4820>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4959(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1561

A DNA sequence (GBSx1653) was identified in *S.agalactiae* <SEQ ID 4821> which encodes the amino acid sequence <SEQ ID 4822>. This protein is predicted to be SNF. Analysis of this protein sequence reveals the following:

-1737-

Possible site: 28

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.32 Transmembrane 743 - 759 (743 - 759)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.1128(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA67095 GB:X98455 SNF [Bacillus cereus]

Identities = 259/678 (38%), Positives = 406/678 (59%), Gaps = 21/678 (3%)

15 Query: 369 QNEILLQMVFDYGNDLTVHNROELEQLTFASHFKHEEKVFKLLEKYGFAPHFSTSHPEAYS 428
 +N +L + F YGN + ++ + F K E+++ ++ + FA + ++
 Sbjct: 388 KNRLLAGLEFHYGNVNVINPLEEDGQPSVFNDRKKEKEILDIMSESAPAKT-EGGYFMHN 446

20 Query: 429 AQELYDFYTYMLPQFKMGTV--SLSAKLESYRLIERPQIDIEAKGSL--LDISFDFSDL 484
 + Y+F +++P K + + + + KL ++ P I + K + L FD +
 Sbjct: 447 EEAENFLYHIVPTLKGVLVDIYATTAKLRIHKGDTAPLIRVRKERIDWLSFRFDIKGI 506

25 Query: 485 LENDVDQALVALFDNNPYFVNKSGQLVIFD-EETKKVSATLQ--GLRARRAKNGHIELDN 541
 E ++ L AL + Y+ +G L+ + +E +++ ++ G+R + +
 Sbjct: 507 PEAEIKGVLAALBEKRYRRLANGSLLSLESKEFNEINQFVKESGIRKEFLHGEEVNVPL 566

30 Query: 542 IAAFQLSELFANQDNVSFSQHFYQLIEDLRHPEKFK--IPGLSVSASLRDYQLTGVRWLS 599
 I + + + +S + L+E +++P+K K +P ++ A +R+YQ+ G W+
 Sbjct: 567 IRSVKWMNGLHEGNVLSLDESVDQLVESIQNPKLKFVTPP-TLHAVMREYQVYGFWMK 625

35 Query: 600 MLDHYGFAGILADDMLGKTLQITISFLSTKLT--RDSR--VLILSPSSLIYNWQDEFHKEF 655
 L +Y F GILADDMLGKTLQ+I+++ + L R+ + +L++SPSSL+YNW E KF
 Sbjct: 626 TLAYYRFGGILADDMLGKTLQSIAYIDSVLPEIREKKLPILVVSPSSLVYNWFSELKGF 685

40 Query: 656 APDVDVAVAYGSKIRREIIAE--RHQVIITSYSSFRQDFETYSEGNYDYLLDEAQVMK 713
 AP + +A G++ R +I+ + V+ITSY R+D +Y+ + L LDEAQ K
 Sbjct: 686 APHIRAVIADGNQTERRKILKDVAEFDVVITSYPLLRDRVRSYARP-FHTLFLDEAQAFK 744

45 Query: 714 NAQTKIAHSLRSFEVKNCFALSGTPIENKLEIWSIFQIILPGLPGKKEFLKLNPKQVA 773
 N T+ A ++++ + + F L+GTP+EN L E+WSIF ++ P LLPG+KEF L + +A
 Sbjct: 745 NPPTQTARAVKTIQAEYRFGLTGTPVENSLEELWSIFHVVPPELLPGRKEFGDLRREDIA 804

50 Query: 774 RYIKPFVMRRRKEEVLPELPLDIEMNYPNEMTDSQKVIYLAQLRQI-QESIQHSSDADLN 832
 +KPFV+RR KE+VL ELPD IE +E+ QK +Y A L ++ +E+++H L
 Sbjct: 805 NAVKPFVLRRLKEDVLQELPDKIEHLOSSELLPDQKRLYAAYLAKLREETLKHLDKDITLR 864

55 Query: 833 RRKIEILSGITRLRQICDTPRLFMD-YDGESGKLESRLQLLTQIKENGHRALIFSQFRGM 891
 + KI IL+G+TRLRQIC+ P LF+D Y G S KLE L +L + + G R LIFSQF M
 Sbjct: 865 KNKIRILAGLTRLRQICNHPALFVDYKSSAKLEQLLDILEECRSTGKRILIFSQFTKM 924

60 Query: 892 LDIAEREMVAMGLTTYKITGSTPANERHEMTRAFNAGSKDAFLISLKAGGVGLNLTGADT 951
 L I RE+ + + +G+TP+ ER E+ FN G D FLISLKAGG GLNLTGADT
 Sbjct: 925 LSIIGRELNRQAIPYFYLDGNTPSQERVELCNRFNEGEGDLFLISLKAGGTGLNLTGADT 984

Query: 952 VVLIDLWNNPAVEMQAISRAHRLGQKENVEVYRLITRGTIIEKILEMQETKKHLVTTVLD 1011
 V+L DLWNNPAVE QA RA+R+GQK V+V +L+ GTIEEK+ E+QE+KKHL+ V++
 Sbjct: 985 VILYDLWNNPAVEQQAADRAYRMGQKNTVQVIKLVAGHTIEEKMHELQESKKHLIAEVIE 1044

Query: 1012 -GNETHASMSVDDIREIL 1028
 G E +S++ ++IR+IL
 Sbjct: 1045 PGEEKLSSITEEIRIDIL 1062

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4823> which encodes the amino acid sequence <SEQ ID 4824>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

65

-1738-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3909(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 674/1031 (65%), Positives = 834/1031 (80%), Gaps = 2/1031 (0%)

10 Query: 1 MSRMIPGRIRNQGIELYEQGLVSLISQEGNLLKAKVGDCQIEYSLVTEETKSCDFFARK 60
 M+R+IPGR+RN+GI+LYEQGLVS +L+ +V Q++Y E+ C CD F K
 Sbjct: 2 MARLIPGRVRNEGILKYEQGLVSFQDDNKGILQIEVETVQVQYGADEEDITCQCDTFHMK 61

15 Query: 61 GYCQHLAALEHFLKNDPEGKAILSKVQVQESQOETKKKTSFGSVFLDSLIIINEDDTIKY 120
 YC+H+AA+E+FLKND +GK L ++ Q + ++ TTK TSFGS+FLDSL +NEDD++KY
 Sbjct: 62 HYCKHIAAVEYFLKNDQKGLFLKQLTNQTKIKETTKMTSFGSLFLDSLAMNEDDSVKY 121

20 Query: 121 QLSAQGEQNPYANDIWWTLKIRRLPDDRSYVIRDIKAFNLTVRKEAYYQIGKQYFETLSL 180
 +LSA G ++P+++D WW+LKI RLPDDRSYVIRDIK FL ++KE +YQIGK YFE LS
 Sbjct: 122 RLSALGSRSPFSSDYWWSLKINRLPDDRSYVIRDIKGFLLQIKKEGFYQIGKNYFEQLSW 181

25 Query: 181 IQFDETSQELIEFLWRLIPSHSSKIDLEFILPNQGRHLSLTRGFFEEGVTLMALENFSF 240
 +QFD +SQ LIEFLWRL S + K D E I PN RHL L GFFEEG+ + +L +F+F
 Sbjct: 182 LQFDPSSQALIEFLWRLA-SDTDKGDNENIFPNHARHLRLPSGFEEGIIHYLTSLYDFTF 240

30 Query: 241 ESDFHQFNHLYFKELEGEDHLYQFKVIVHRQSIELEIKEKDLKPLFANSYLFYRDTFYHL 300
 E ++HL+ + LE E LY+FKV VHR+SIEL+I EK+++ LF N YL Y+DTFYHL
 Sbjct: 241 EGPSTQTYHHLFVRSLEAAGLYEFKVEVHRKSIELQIAEKNVQYLFNDYLLYQDTFYHL 300

35 Query: 301 NLKQEKMTAIRSLPIEGDLAKHIHFDLDDQDKLAHLLDFKEIGLVDAPRSFSIHDFKV 360
 LKQ KMV AIRSLPIE DLAKHIHFDLDD KLAA L DFK+IGLV+AP+SF+I DF+V
 Sbjct: 301 TLKQKRMVQAIRSLPIEADLAKHIHFDLDDHAKLAASLSDFKQIGLVEAPKSF AIRDFEV 360

40 Query: 361 NFEFDINSQNEILLQMFVFDYCNLDLTVHNRQELEQLTFASHFKHEEKVFKLLEKYGFAPHF 420
 F+FD+ +++EI Q++FDYGN V ++ LE L FASH K EEK+ + L +GF+P F
 Sbjct: 361 TFQFDLLNRDEISQCLMFYDGN-YQVSDKASLEALPFASHLKKEEKINRSLAFGFSFQF 419

45 Query: 421 STSHPAYSAQELYDFYTYMLPQFKMGTVSLSAKLESYRLIERPQIDIEAKGSLDISFD 480
 + SA+ELY F+ +P F+++G V+LS +++ ++ E P+I I LLDISFD
 Sbjct: 420 YSKRLTSAKELYTFEETVPCFERLGNVALSTAIQALQVKEMPKIAIRRNQGLDISFD 479

50 Query: 481 FSDLLENDVDQALVALFDNNPYFVNKSGQLVIFDEETKKSATLQGLRARRAKNGHIELD 540
 FS ++END+DQA+ ALF NNPYFV+++GQLV+FD+ET+KVS +LQ LRAR+ KNGH++LD
 Sbjct: 480 FSTIIENDIDQAVTALFQNNPYFVSQTGQLVVFDEETQKVSLSLQELRARQLKNGHLQLD 539

55 Query: 541 NIAAFQLSLNFANQDNVSFSQHLYQLIEDLRHPEKFKIPGLSVSASLRDYQLTGVRWLSM 600
 I A Q+S+LF +V FS+ +L L+HPE F I L V A +RDYQ GV+WLSM
 Sbjct: 540 GIRALQVSKLFEGMTSVHFSKELELAYHLQHPETFSIKPLPVKAQMRDYQRNGVQWLSM 599

60 Query: 601 LDHYGFAGILADDMGLGKTLQTLISFLSTKLTRDSRVLLILSPSSLIYNWQDEFHKFAPDVD 660
 L+HYGF GILADDMGLGKTLQTL++FL++ L DS+VLILSPSSLIYNW DE KF P +D
 Sbjct: 600 LNHYGFGILADDMGLGKTLQTLAFLASHLKS DSKVLILSPSSLIYNWFDECCQKFTPQLD 659

65 Query: 661 VAVAYGSKIRREIIAERHQVIITSYSSFRQDFETYSEGNYDYLLILDEAQVMKNAQTKIA 720
 V V+YG K RD+II E HQ+ ITSYSSFRQDFETY +YDYLLILDEAQV+KNAQTKI+
 Sbjct: 660 VVVSGLKQIRDQIEEGHQITITSYSSFRQDFETYQAFHYDYLLILDEAQVIKNAQTKIS 719

Query: 721 HSLRSFEVKNCFALSGTPIENKLEIWSIFQIILPGLLPKKEFLKLNPKQVARYIKPFV 780
 H LR+F NCFALSGTPIENK+LEIWSIFQI+LPGLLP KKEFLKL +QV+RYIKPFV
 Sbjct: 720 HCLRAFNTANCFALSGTPIENKMLEIWSIFQIVLPGLLPKKEFLKLTAEQVRSYIKPFV 779

Query: 781 MRRRKEEVLPELPLDIEMNYPNEMTDSQKVIYLAQLRQIESIQHSSDADLNRRKIEILS 840
 MRR+KE+VLPPLDLIE+NY NEMTD QK IYLAQLRQ+Q+ I++SSD D++R+KIEILS
 Sbjct: 780 MRRKEDVLPPLDLIEINYSNEMTDEQKAIYLAQLRQMQDQIRNSSDVIDSRQKIEILS 839

65 Query: 841 GITRLRQICDTPRLFM DYD GSGKLESRLQLLTQIKENGHRALIFSQFRGMLDIAEREMV 900
 GITRLRQICDTP LFM DY G+SGL+SLR LLTQIKENGHRALIFSQFRGMLD+A++EM

-1739-

Sbjct: 840 GITRLRQICDTPSLFMDYQGKSGKLDLSRILLTQIKENGHRALIFSQFRGMLDLAKQEMT 899
 Query: 901 AMGLITTYKITGSTPANERHEMTRAFNAGSKDAFLISLKAGGVGLNLTGADTVVLIDLWWN 960
 A+GLT+Y++TGSTPANER EMTRAFN GSKDAFLISLKAGGVG+NLTGADTV+LIDLWWN
 5 Sbjct: 900 ALGLTSYQMTGSTPANERQEMTRAFNNGSKDAFLISLKAGGVGINLTGADTVILIDLWWN 959
 Query: 961 PAVEMQAISRAHRLGQKENVEVYRLITRGTIIEKILEMQETKKHLVTTVLDGNETHASMS 1020
 PAVEMQAISRA+R+GQKENVEVYRLITRGTIIEKILE+QE+K++LVTTVLDGNE+ ASMS
 10 Sbjct: 960 PAVEMQAISRAYRIGQKENVEVYRLITRGTIIEKILELQESKRNLVTTVLDGNESRASMS 1019
 Query: 1021 VDDIREILGVS 1031
 +++I+EILG++
 Sbjct: 1020 IEEIKEILGLN 1030

15 SEQ ID 4822 (GBS369) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 5; MW 120kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 6; MW 142kDa).

The GBS369-GST fusion product was purified (Figure 215, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 303), which confirmed that the protein is immunoaccessible
 20 on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1562

A DNA sequence (GBSx1654) was identified in *S.agalactiae* <SEQ ID 4825> which encodes the amino
 25 acid sequence <SEQ ID 4826>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.3391(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 There is also homology to SEQ ID 1034:

Identities = 34/38 (89%), Positives = 37/38 (96%)
 Query: 1 MEKEAKQIIDLKRNLFKIDVRAQKDEEKVFMRTACQFS 38
 +EKEAKQ+IDLKRNLFKIDVRAQKDEEKVFMRTAC+ S
 40 Sbjct: 1 LEKEAKQMIDLKRNLFKIDVRAQKDEEKVFMRTACRQS 38

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1563

45 A DNA sequence (GBSx1656) was identified in *S.agalactiae* <SEQ ID 4827> which encodes the amino acid sequence <SEQ ID 4828>. This protein is predicted to be phosphoglycerate dehydrogenase (era2). Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence
 50

-1740-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3709(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA88823 GB:AB016077 phosphoglycerate dehydrogenase

[Streptococcus mutans]

Identities = 377/436 (86%), Positives = 414/436 (94%)

Query: 1 MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIYTTGEWLNKRFSLIDTGG 60
 M LPTVAIVGRPNVGKS LFNRIAGERISIVEDVEGVTRDRIYT EWLNR+FS+IDTGG
 Sbjct: 1 MALPTVAIVGRPNVGKSALFNRIAGERISIVEDVEGVTRDRIYTKAEWLNRFSLIDTGG 60

Query: 61 IDVDVAPFMEQIKHQADIAMTEADVIVFVSGKEGVTDADEYVSRIYKTNKPVILAVNK 120
 IDVDVAPFMEQIKHQADIAMTEADVIVFVVS KEG+TDADEYV++ILY+T+KPVILAVNK
 Sbjct: 61 IDVDVAPFMEQIKHQADIAMTEADVIVFVVSKEGITDADEYVAKILYRTHKPVILAVNK 120

Query: 121 VDNPEMRNDIYDFYSLGLGDPYPLSSVHGIGTGDILDAIVENLPVEEENENPDIIIRFSLI 180
 VDNPEMR+ IYDFY+LGLGDPYP+SS HGIGTGD+LDAIV+NLP E + E+ DII+FSLI
 Sbjct: 121 VDNPEMRSAIYDFYALGLGDPYPVSSAHGIGTGDVLDIVDNLPAAEQEESDIIKFSLI 180

Query: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTNFVDSQGEYTMIDTAGMRKSGKVY 240
 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDT F D +GQE+TMIDTAGMRKSGKVY
 Sbjct: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTTFDEEGQEFTMIDTAGMRKSGKVY 240

Query: 241 ENTEKYSVMRSMRAIDRSADVLMVINAEEGIREYDKRIAGFAHETGKIIIVVNKWDITIE 300
 ENTEKYSVMR+MRAIDRS+VLMV+NAEEGIREYDKRIAGFAHE GKGI++VVKWD I+
 Sbjct: 241 ENTEKYSVMRAMRAIDRSDIVLMVINAEEGIREYDKRIAGFAHEAGKGI VVVVNKWDIAIK 300

Query: 301 KDNHTVSQWEADIRDNFQFLSYAPIIFVSAETKQRLHKLDPDMIKRISQNKRIIPSAVLN 360
 KDN TV+QWE DIRDNFQ++ YAPI+FVSA TKQRLHKLDP+IK++S+SQN RIPS+VLN
 Sbjct: 301 KDNRTVAQWETDIRDNFQYIPYAPIFVSAVTKQRLHKLDPVIKQVSQSNTRIPSSVLN 360

Query: 361 DVIMDAIAINPTPTDKGKRLKIFYATQVAVKPPTFVVFVNEEELMHFSYLRFLENQIREA 420
 DV+MDA+AINPTPTDKGKRLKIFYATQV+VKPPTFV+VNEEELMHFSYLRFLENQIR+A
 Sbjct: 361 DVVMDAVAINPTPTDKGKRLKIFYATQVSVKPPTFVIFVNEEELMHFSYLRFLENQIRQA 420

Query: 421 FVFEETPINLIARKRK 436

FVFEETPI LIARKRK

Sbjct: 421 FVFEETPIRLIARKRK 436

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4829> which encodes the amino acid sequence <SEQ ID 4830>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3463(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 403/436 (92%), Positives = 422/436 (96%)

Query: 1 MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIYTTGEWLNKRFSLIDTGG 60
 MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIY TGEWLNRFSLIDTGG
 Sbjct: 1 MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIYATGEWLNRFSLIDTGG 60

Query: 61 IDVDVAPFMEQIKHQADIAMTEADVIVFVSGKEGVTDADEYVSRIYKTNKPVILAVNK 120
 IDVDVAPFMEQIKHQ IA EADVIVFVSGKEGVTDADEYVS+ILY+TN PVILAVNK
 Sbjct: 61 IDVDVAPFMEQIKHQAQIAMEEADVIVFVSGKEGVTDADEYVSKILYRTNTPVILAVNK 120

Query: 121 VDNPEMRNDIYDFYSLGLGDPYPLSSVHGIGTGDILDAIVENLPVEEENENPDIIIRFSLI 180

-1741-

VDNPEMRNDIYDFYSLGLGDPYP+SSVHGIGTGD+LDAIVENLPVEE EN DIIRFSLI
 Sbjct: 121 VDNPEMRNDIYDFYSLGLGDPYPVSSVHGIGTGDVLDIAIVENLPVEEAEENDDIIRFSLI 180

5 Query: 181 GRPNVGKSSSLINAILGEDRVIASPVAGTTTRDAIDTNFVDSQGQEYTMIDTAGMRKSGKVY 240
 GRPNVGKSSSLINAILGEDRVIASPVAGTTTRDAIDT+F D+ GQE+TMIDTAGMRKSGK+Y
 Sbjct: 181 GRPNVGKSSSLINAILGEDRVIASPVAGTTTRDAIDTHTFDADGQEFMIDTAGMRKSGKIY 240

10 Query: 241 ENTEKYSVMRSMRAIDRSDDVLMVINAEEGIREYDKRIAGFAHETGKGIIIVVNKWDITIE 300
 ENTEKYSVMR+MRAIDRSDDVLMVINAEEGIREYDKRIAGFAHE GKG+IIVVNKWDITI+
 Sbjct: 241 ENTEKYSVMRAMRAIDRSDDVLMVINAEEGIREYDKRIAGFAHEAGKGMIIIVVNKWDITID 300

15 Query: 301 KDNHTVSQWEADIRDNFQFLSYAPIIFVSAETKQRLHKLPMIKRISSEONKRIPSAVLN 360
 KDNHTV++WEADIRD FQFL+YAPIIFVSA TKQRL+KLDP+IKRISSEONKRIPSAVLN
 Sbjct: 301 KDNHTVAKWEADIRDQFQFLTYAPIIFVSALTQRLNKLPLIKRISSEONKRIPSAVLN 360

20 Query: 361 DVIMDAIAINPTPTDKGKRLKIFYATQVAVKPPTFVVFNNEELMHFSYLRFLENQIRAA 420
 DVIMDAIAINPTPTDKGKRLKIFYATQV+VKPPTFVVFNNEELMHFSYLRFLENQIR A
 Sbjct: 361 DVIMDAIAINPTPTDKGKRLKIFYATQVSVKPPTFVVFNNEELMHFSYLRFLENQIRAA 420

Query: 421 FVFEGTPIINLIARKRK 436
 F FEGTPI+LIARKRK
 Sbjct: 421 FTFEGTPIHILIARKRK 436

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 1564

A DNA sequence (GBSx1657) was identified in *S. agalactiae* <SEQ ID 4831> which encodes the amino acid sequence <SEQ ID 4832>. Analysis of this protein sequence reveals the following:

Possible site: 51
 30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2734(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00359 GB:AF008220 DnaI [Bacillus subtilis]
 Identities = 105/313 (33%), Positives = 191/313 (60%), Gaps = 17/313 (5%)

40 Query: 1 MKSVGQALENQGRVP--RNTNDELIQMILADAQVAEFIKTHQ--LSQREINISMSKFNQF 56
 M+ +G++L+ P + +++ + ++ D V F+K ++ + Q+ I S++K ++
 Sbjct: 1 MEPIGRSLQGVGTGRPDFQKRLEQMKEKVMKDQDVQAFLENEEVIDQKMIKSLNKLIEY 60

45 Query: 57 LIERQK-----FKNKDSQYIAGGYEPILVMNEGYADVSYLE--TRELIEAQKKQAISDRI 109
 IE+ K +++ + +GY P LV+N D+ Y E + ++ QKKQ +
 Sbjct: 61 -IEQSKNCSYCEDENCNNLLEGYHFKLVNCRSIDIEYYECPVKRKLDDQKKQ--QSLM 117

50 Query: 110 NLVNLPKSYRNIRMTDFDINNESRMKAMSQLLDFVETYPYNYH-KGLYLYGDMGVGKSYL 168
 + + + DI++ SR+ + DF+++Y KGLYLYG GVGK+++
 Sbjct: 118 KSMYIQDQLLGATFQQVDISDPSRLAMFQHVTDFLKSYNETGKGLYLYGKFGVGKTFM 177

55 Query: 169 MAAMARELSERKGVSTTLHFPSPFAIDVKNAISSGTVKDEIDAVKSVPIILDDIGAEQA 228
 +AA+A EL+E++ S+ +++ P F ++KN++ T+++++ VK+ P+L+LDDIGAE
 Sbjct: 178 LAAIANELABKE-YSSMIVYVPEFVRELKNSLQDQTLEEKLNVMVKTTPVLMMLDDIGAESM 236

60 Query: 229 TSWVRDEILQVILQHRMLEELPTFFTSNYSFNDLERKWA-NIKGSDETQAKRVMERVRY 287
 TSWVRDE++ +LQHRM ++LPTFF+SN+S ++L+ + + +G E +A R+MER+ Y
 Sbjct: 237 TSWVRDEVIGTVLQHRMSQQLPTFFSSNFSPELKHFTYSQRGEKEEVKAARLMERILY 296

Query: 288 LAIEFHLEGNRR 300
 LA L+G NRR

-1742-

Sbjct: 297 LAAPIRLDGENRR 309

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4833> which encodes the amino acid sequence <SEQ ID 4834>. Analysis of this protein sequence reveals the following:

5 Possible site: 19
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1944(Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 228/300 (76%), Positives = 264/300 (88%)

15 Query: 1 MKSVGQALENQGRVPRNTNDELIQMILADAQVAEFIKTHQLSQREINISMSKFNQFLIER 60
 M+ +G+ + G+ R +D+LIQ ILAD +VA FI H LSQ +IN+S+SKFNQFL+ER
 Sbjct: 1 MEKIGETMAKLGQNTRVNSDQLIQITILADPEVASFISQHHLSQEQINLSLSKFNQFLVER 60

20 Query: 61 QKFKNKDSQYIAKGYEPILVMNEGYADVSYLETRELIEAQKKQAISDRINLVNLPKSYRN 120
 QK++ KD YIAKGY+PIL MNEGYADVSYLET+EL+EAQK+ AIS+RI LV+LPKSYR+
 Sbjct: 61 QKYQLKDPSTYIAKGYQPILAMNEGYADVSYLETKELVEAQKQAISERIQLVSLPKSYRH 120

25 Query: 121 IRMTDFDINNESRMKAMSQQLLDFVETYPSTYHNKGLYLYGDMGVGKSYLMAAMARELSERK 180
 I ++D D+NN SRM+A S +LDFVE YPS KGLNLYGDMG+GKSYL+AAMA ELSE+K
 Sbjct: 121 IHLSDIDVNNASRMEAFSAILDFVEQYPSAEQKGLYLYGDMGIGKSYLLAAMAHELSEKK 180

30 Query: 181 GVSTTLHFPSPFAIDVKNAISSGTVKDEIDAVKSVPIILDDIGAEQATSWVRDEILQVI 240
 GVSTTLHFPSPFAIDVKNAIS+G+VK+EIDAVK+VP+LILDDIGAEQATSWVRDE+LQVI
 Sbjct: 181 GVSTTLHFPSPFAIDVKNAISNGSVKEEIDAVKNVPVILDDIGAEQATSWVRDEVILQVI 240

35 Query: 241 LQHRMLEELPTFFTSNYSFNLERKWKANIKGSDETWQAKRVMERVRYLAIEFHLEGPNRR 300
 LQ+RMLEELPTFFTSNYSF DLERKWA IKGSDETWQAKRVMERVRYLA EFHLEG NRR
 Sbjct: 241 LQYRMLEELPTFFTSNYSFADLERKWATIKGSDETWQAKRVMERVRYLAREFHLEGANRR 300

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1565

40 A DNA sequence (GBSx1658) was identified in *S.agalactiae* <SEQ ID 4835> which encodes the amino acid sequence <SEQ ID 4836>. Analysis of this protein sequence reveals the following:

 Possible site: 37
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2660(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4837> which encodes the amino acid sequence <SEQ ID 4838>. Analysis of this protein sequence reveals the following:

 Possible site: 35
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2135(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1743-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 217/391 (55%), Positives = 309/391 (78%)

5 Query: 1 MMSPIDEFTYIKQNKIVYDSNSLIQLYFPIMGSDAMALYDYFVHFDDGIRRHKFSEVLN 60
MM PID FTY+K+NK+ DS +LIQLYFPI+GSDA+++Y YF+HFFDDG++RHKFS++LN
Sbjct: 1 MMKPIDTFTYLKRNVTLDSVTILQLYFPIIGSDAVSIYQYFIHFFDDGLQRHKFSDIILN 60

10 Query: 61 HLQYGMPRFQDALVMLTALDLLTVYQATGTYLVLKLNQAMSNEFLSNPIYRRLLEKRIGE 120
HLQ+GM RF+DAL +LTA++L++VYQ + TYL+ L+Q +S +LF +P Y RLLE++IGE
Sbjct: 61 HLQFGMKRFEDALAILTAMELVSVYQLSDTYLITLHQPLSRDLFFQHPAYSRLLEQKIGE 120

15 Query: 121 VAVAEELDMKIPKNARDISKFTDVFSDLGQPKQEVNRSKNVFDLESFKRLMMRDGLRFNN 180
VAV+EL + +P AR+ISK+F+D+F G + + FDL SF++LM+RDGL+F +
Sbjct: 121 VAVSELQVTVPSQARNISKRFSDIFGVQGDLTNVPQKPQKNFDLSSFQQLMVRDGLQFED 180

20 Query: 181 EKDDVLGIYSVSELYHLNWDYTYQLAKQTAINGMIAPQRMKVQQNEGQHIKDNQSFTNNE 240
+ D++ +YS++E Y + W+DTYQ+AK TA+NG I P+R+ ++N+ ++F+ E
Sbjct: 181 NQKDIISLYSLAEQYDMTWFDYQIAKATAVNGKIRPERLLAKKNQSMTKPSKENFSQAE 240

25 Query: 241 KVILRESKNSALVFLEKIKRSRKAVTTSSEKTLLEDLAKMNFLEVINVMVLYTLNKT 300
++ILRE+K DSALVFLEKIK++R+A T E+ LL+ LAKMNFLE+VINVMVLYT NKT
Sbjct: 241 QIILREAKQDSALVFLEKIKARRATITKDERILLQLAKMNFLEDDVINVMVLYTFNKT 300

30 Query: 301 SANLNKAYIMKVANDFAFQNVMTAEADAVLKIRDFSDQKVRTKTETKKQSNVPEWSNP 360
SANL K+Y++K+ANDFA+Q V TAE+AA+ +R F+D++ R +++ K QSNVP+WSNP
Sbjct: 301 SANLQKSYVLKMANDFAYQKVSTAEAAIVVLRAFTDRQSRQSKVKTQSQSNVPKWSNP 360

30 Query: 361 KDEVSPEKEIELEQFKTDALKRLERLGKGE 391
++ S E++ +L+QFK ALKRL E LGK G+
Sbjct: 361 QETTSQEEQAKLDQFKQAALKRLNLGKGGD 391

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1566

A DNA sequence (GBSx1659) was identified in *S. agalactiae* <SEQ ID 4839> which encodes the amino acid sequence <SEQ ID 4840>. Analysis of this protein sequence reveals the following:

Possible site: 19
40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4485(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06865 GB:AP001517 unknown conserved protein [Bacillus halodurans]
Identities = 80/150 (53%), Positives = 115/150 (76%)

50 Query: 1 MRCPKCGYNKSSVDSRQAEEGTTIRRRRECEKCGNRFTTTFERLEELPLLVIKKGDTREQ 60
MRCP C +N + V+DSR A EG +IRRRRECE C +RFTTTE +EE+PL+V+KKGDT++
Sbjct: 1 MRCPACHHNGTRVLDSPAEHGRSIRRRRECECNHRTTTFEMIEEVPLIVVKKGDTREQ 60

55 Query: 61 FSRDKILNGIIQSAQKRPVSSEDIENCILRIERKIRSEYEDVSSITIGNLVMDLAELD 120
FS DKIL G+I++ +KRPV E +E + +ER++R + ++EV S IG LVM+ LA +D
Sbjct: 61 FSSDKILRLIRACEKRPVPLETLEGIVNEVERELRGQKNEVDSKEIGELVMERLANVD 120

60 Query: 121 EITYVRFASVYKSFKDVEIEELLQQITKR 150
++ YVRFASVY+ FKD++ + L+++ +R
Sbjct: 121 DVAYVRFASVYRQFKDINVFIQELKELMER 150

-1744-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4841> which encodes the amino acid sequence <SEQ ID 4842>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4365(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 131/155 (84%), Positives = 143/155 (91%)

Query: 1 MRCPKCGYNKSSVDSRQAEAGTTIRRRRECEKCGNRFTTFFERLEELPLLVIKKGDTREQ 60
 +RCPKC Y+KSSVDSRQAE+G TIRRRRECE+C RFTTFFER+EELPLLVIKKGDTREQ
 Sbjct: 1 VRCPKCNYHKSSVDSRQAEAGNTIRRRRECEQCHTRFTTFFERVEELPLLVIKKGDTREQ 60

Query: 61 FSRDKILNGIIQSAQKRPVSSEDIENCILRIERKIRSEYEDEVSSITIGNLVMDLAELD 120
 FSRDKILNG++QSAQKRPVSS DIEN I RIE+++R+ YE+EVSS IGNLVMDLAELD
 Sbjct: 61 FSRDKILNGVVQSAQKRPVSSDTIENVISRIEQEVRTTYENEVSSTAIGNLVMDLAELD 120

Query: 121 EITYVRFASVYKSFKDVEIEELLQQTITKRVRSKK 155
 EITYVRFASVYKSFKDVEIEELLQQTIT RVR KK
 Sbjct: 121 EITYVRFASVYKSFKDVEIEELLQQTITNRVRGKK 155

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1567

A DNA sequence (GBSx1660) was identified in *S.galactiae* <SEQ ID 4843> which encodes the amino acid sequence <SEQ ID 4844>. This protein is predicted to be CsrS (mtrB). Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -11.30 Transmembrane 22 - 38 (18 - 43)
 INTEGRAL Likelihood = -9.66 Transmembrane 189 - 205 (187 - 212)

----- Final Results -----
 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2109> which encodes the amino acid sequence <SEQ ID 2110>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -6.32 Transmembrane 196 - 212 (189 - 214)

----- Final Results -----
 bacterial membrane --- Certainty=0.3527(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 248/501 (49%), Positives = 363/501 (71%), Gaps = 4/501 (0%)

-1745-

Query: 1 MKNKKDQFIGVKQPLSKKLSQLVFLFFSLFTVFSVLVYTSATRYVLHREKINVGRSLEK 60
 M+N+K + K L K+LS + F+LFF +F+ F+++ Y+S ++L +EK +V +++
 Sbjct: 1 MENQKQKQKKYKNSLPKRLSNIFFVLFFCIFSFTLIAYSSSTNYFLKKKEQSVFQAVNI 60

5 Query: 61 TRVRLSQANSSLTSDILEILYNQVFADDIYPHKRQNGIVRTGESIDSILYVNQEMTLVD 120
 RVRLS+ +S+ T +++ E+LY ++ + ++R+ I + L NQ++ +Y+
 Sbjct: 61 VRVRLSEVDSNFTLENLAELVLYKNDKTHLRIDDRKGSRVIRSERDITNTLDANQDIYVYN 120

10 Query: 121 VNRKPVFST-LRTGMPITIGKSMGKVIISKVADM-EGFVGTKAIYSQKTGQLLGVQIFYN 178
 ++++ +F+T P + +G+V + D GF T+ +YS +TG+ +GYVQ+F++
 Sbjct: 121 IDKQMIFTTDNEESSPGLHGPGRVYHDHIEDQYRGFSMTQKVYSNRTGKFGVGVQVPHD 180

15 Query: 179 LGRYYSMRQNIIVFLIMMEVLGTVLALVVINSATKRIVRPVKNLHDLMHQISENPSNLEI 238
 LG YY +R ++ +L+++E+ GT LA ++I T+R ++P+ NLH++M ISENPNL +
 Sbjct: 181 LGNYVIRARLLFWLLVVELFGTSLAYLILITRRFLKPLHNLHEVMRNISENPNNLNL 240

20 Query: 239 RSKVRSEDEIGELSRIFDGMLEDOEDYTRRQSQFISDVSHELRTPVAVVKGHIGLLQRWG 298
 RS + S DEI ELS IFD MLD+LE +T+ QS+FISDVSHELRTPVA++KGHIGLLQRWG
 Sbjct: 241 RSDISSGDEIEELSVIFDNMLDKLEHTTKLQSRFISDVSHELRTPVAIIKGHIGLLQRWG 300

25 Query: 299 KDDPEILLESIAAYHEADRMSIMINDMLNMIRVQGSLELHQDEVTDLSSSISVVIENFR 358
 KDD +ILEESL A HEADRM++MINDML+MIRVQGS E HQ+++T L SI V+ NFR
 Sbjct: 301 KDDSDILEESLTATAHEADRMAIMINDMLDMIRVQGSFEGHQNDMTVLEDSIETVVGNER 360

30 Query: 359 ILREDFQFIFENNISDIVWGKIYKIHFEQALMILIDNAIKYSPSYKEVSVVLSVDNDFAT 418
 +LREDF F +++ + +IYK HFEQALMILIDNA+KYS K++++ LSV
 Sbjct: 361 VLREDFIFTWQSENPKTI-ARIYKNHFEQALMILIDNAVKYSRKEKKIAINLSVTGKQEA 419

35 Query: 419 VV-VKDKGEGISDEIDIEFIFDRFYRTDKSRNRESTQAGLGIGLSVFKQIMDAYHLKVDIK 477
 +V V-DKGEIGIS EDIE IF+RFYRTDKSRNR STQAGLGIGLS+ KQI+D YHL++ ++
 Sbjct: 420 IVRVQDKGEGISKEDIEHIFERFYRTDKSRNRTSTQAGLGIGLSILKQIVDGYHLQMKVE 479

Query: 478 SELNQCTEFIVRIPIKKFEET 498
 SELN+G+ FI+ IP+ + +E+
 Sbjct: 480 SELNEGSVFILHIPLAQSKES 500

A related GBS gene <SEQ ID 8845> and protein <SEQ ID 8846> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop: Possible site: -1 Crend: 5
 SRCFLG: 0
 McG: Length of UR: 5
 Peak Value of UR: 0.74
 Net Charge of CR: 2
 McG: Discrim Score: -10.19
 45 GvH: Signal Score (-7.5): -3.66
 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 Amino Acid Composition: calculated from 1
 ALOM program count: 2 value: -11.30 threshold: 0.0
 50 INTEGRAL Likelihood = -11.30 Transmembrane 22 - 38 (18 - 43)
 INTEGRAL Likelihood = -9.66 Transmembrane 189 - 205 (187 - 212)
 PERIPHERAL Likelihood = 2.86 405
 modified ALOM score: 2.76
 icml HYPID: 7 CFP: 0.552
 55 *** Reasoning Step: 3
 ----- Final Results -----
 60 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1746-

SEQ ID 8846 (GBS321) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 6; MW 84kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 2; MW 58.7kDa).

GBS321-GST was purified as shown in Figure 220, lane 3.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1568

- A DNA sequence (GBSx1661) was identified in *S.agalactiae* <SEQ ID 4845> which encodes the amino acid sequence <SEQ ID 4846>. This protein is predicted to be CsrR (trcR). Analysis of this protein
10 sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2649(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3259> which encodes the amino acid
20 sequence <SEQ ID 3260>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3226(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 30 Identities = 193/229 (84%), Positives = 211/229 (91%), Gaps = 1/229 (0%)
Query: 1 MGKKILIIIEDEKNLARFVSLELLHEGYDVVVEITNGREGLDTALEKDFDLILLDLMLPEMD 60
M KKILIIIEDEKNLARFVSLEL HEGY+V+VE NGREGL+TALEK+FDLILLDLMLPEMD
Sbjct: 1 MTKKILIIIEDEKNLARFVSLELQHEGYEVIVEVNGREGLETALEKEFDLILLDLMLPEMD 60
35 Query: 61 GFEITRRLQAEKTTYIMMTARDSVMDIVAGLDRGADDYIVKPF AIEELLARVRAIFRRQ 120
GFE+TRRLQ EKTYYIMMTARDS+MD+VAGLDRGADDYIVKPF AIEELLAR+RAIFRRQ
Sbjct: 61 GFEVTRRLQTEKTTYIMMTARDSIMDVVAGLDRGADDYIVKPF AIEELLARIRAIFFRRQ 120
40 Query: 121 EIETKTKEKGDGSGSFRDLSLNTHNRSAMRGDEEISLTKREFDLLNVLMTNMNRVMTREEL 180
+IE++ K+ G +RDL LN NRS RGD+EISLTKRE+DLIN+LMTNMNRVMTREEL
Sbjct: 121 DIESE-KKVPSQGIYRDLVLNPQNRSVNRGDDEISLTKREYDLLNLMTNMNRVMTREEL 179
45 Query: 181 LEHVWKYDVAETNVVDVYIRYLRGKIDIPGRESYIQTVRGMGYVIREK 229
L +VWKYD A ETNVVDVYIRYLRGKIDIPG+ESYIQTVRGMGYVIREK
Sbjct: 180 LSNVWKYDEAVETNVVDVYIRYLRGKIDIPGKESYIQTVRGMGYVIREK 228

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 1569

A DNA sequence (GBSx1662) was identified in *S.agalactiae* <SEQ ID 4847> which encodes the amino acid sequence <SEQ ID 4848>. Analysis of this protein sequence reveals the following:

-1747-

Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5 bacterial cytoplasm --- Certainty=0.3864(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 10 >GP:AAG32547 GB:U12643 YlbN-like hypothetical protein [Streptococcus gordonii]
 Identities = 91/174 (52%), Positives = 133/174 (76%), Gaps = 3/174 (1%)

- Query: 3 LTEIKKSPEGLYFDKKIDIKESLMERHSEIMDISDIQVSGHVYEDGLYLDDYNMAYDIT 62
 + EI+K+P+GL F+KK+D+ E L ER++EI+D+ DI SG YEDGLY LDY ++Y IT
 15 Sbjct: 4 IQEIRKNPDGLAFEKKLDLAEELKERNAEILDVQDIVASGRAQYEDGLYFLDYELSYTIT 63
- Query: 63 LPSSSRMKPVVLSEKQTINEVFIEAENVSTKKELVDQELVLILEEDDINLEESVIDNILL 122
 L SSRSM+PV E +NE+F+E V++ +E++DQ+LVL +E +IN+ ESV DNILL
 20 Sbjct: 64 LASSRSMFVERKESYLVNEIFMEDGQVAS-QEMIDQDLVLPIENGAINVAESVADNILL 122
- Query: 123 NIPLRVL-AADEVGV EADLSGKNWSLMTEKQYEEKQAKEKEKSNPFAALEGMFD 175
 NIPL+VL AA+E G + +G++W +MTE Y++ QA++KE+++PFA L+G+FD
 Sbjct: 123 NIPLKVLTAABEAGSDLP-TGRDWQVMTEDDYQKYQAEKKEENSPFAGLQGLFD 175

- 25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4849> which encodes the amino acid
 sequence <SEQ ID 4850>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 30 bacterial cytoplasm --- Certainty=0.3032(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 35 An alignment of the GAS and GBS proteins is shown below.

Identities = 86/175 (49%), Positives = 135/175 (77%)

- Query: 1 MLLTEIKKSPEGLYFDKKIDIKESLMERHSEIMDISDIQVSGHVYEDGLYLDDYNMAYD 60
 + ++EI+K P+GL FD+ D+K L+ER +I+DI ++ G+V Y+ GLYLLDY ++Y+
 40 Sbjct: 3 LAISEIRKHPDGLSFDRLCDVKSMLLERDQQIIDIKAVKAVGNVRYDKGLYLLDYQLSYE 62
- Query: 61 ITLPSSSRMKPVVLSEKQTINEVFIEAENVSTKKELVDQELVLILEEDDINLEESVIDNI 120
 + LPSSSRM PV LSE Q I E+FIEA +++ KKELV+ LVL+L++D INLEES++DNI
 45 Sbjct: 63 VILPSSSRMVPVCLSEVQHIQELFIEATDLADKKELVEDNLVLVLDKDAINLEESIVDNI 122
- Query: 121 LLNIPLRVLAADEVGV EADLSGKNWSLMTEKQYEEKQAKEKEKSNPFAALEGMFD 175
 LL IP++VL +E + +G+NW+++TE+ Y+ + ++++++NPFA+L+G+FD
 Sbjct: 123 LLAIPVQVLTEEEKKSKELPAGQNWAVLTEEDYQCLKEEKQKENNPFAQLGLFD 177

- 50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1570

- A DNA sequence (GBSx1663) was identified in *S.agalactiae* <SEQ ID 4851> which encodes the amino
 acid sequence <SEQ ID 4852>. This protein is predicted to be heat shock protein (htpX). Analysis of this
 55 protein sequence reveals the following:

Possible site: 25

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -11.30 Transmembrane 195 - 211 (190 - 221)

-1748-

INTEGRAL Likelihood = -11.09 Transmembrane 43 - 59 (31 - 62)
 INTEGRAL Likelihood = -3.61 Transmembrane 153 - 169 (153 - 174)

----- Final Results -----

5 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAB70525 GB:AF017421 putative heat shock protein HtpX
 [Streptococcus gordonii]
 Identities = 220/297 (74%), Positives = 261/297 (87%), Gaps = 1/297 (0%)

Query: 1 MLYQQIASNKRKTVVLLIVFFCLLAAIGAAGVYLVLGSYQFGLVLALIIGVIYAVSMIFQ 60
 15 ML++QIA+NKR+T LL+ FF LLA IGAA GYL + S G+++A IIG+IYA++MIFQ
 Sbjct: 1 MLFEQIAANKRRTWFLVAVFALLALIGAAAGYLMWNSPLGGVIIAFIIGLIYAITMIFQ 60

Query: 61 STNVVMSMNAREVTEDEAPNYFHIVEDMAMIAQIPMPRVFIVEDDSLNAFATGSKPEN 120
 ST VVMSMN AR+V+E EAP +HIV+DMAM+AQIPMPRV+IVEDDS NAFATGS PEN 120
 20 Sbjct: 61 STEVVMSMNGARQVSEQEAPELYHIVQDMAMVAQIPMPRVYIVEDDSPNAFATGSPEN 120

Query: 121 AVAATTGLLAVMNRREELEGVIGHEVSHIRNYDIRISTIAVALASAVTLISSIGSRMLFYG 180
 AVAATTGLL +MNRREELEGVIGHEVSHIRNYDIRISTIAVALASA+T+ISS+ RM++YG
 25 Sbjct: 121 AVAATTGLLRLMNRREELEGVIGHEVSHIRNYDIRISTIAVALASAITMISSVAGRMMWYG 180

Query: 181 GRRRRDDDDREDDG-NILVLIFSILSLILAPLAASLVQLAISRQREYLADASSVELTRNPQ 239
 GRRRR-D +D G +L+L+FS++++ILAPLAA+LVQLAISRQRE+LADASSVELTRNPQ
 30 Sbjct: 181 GRRRRNDRDDSGLGLLMLVFSLIAIILAPLAATLVQLAISRQREFLADASSVELTRNPQ 240

Query: 240 GMISALEKLD RSEPMGHPVDDASAALYINDPTKKEGLKSLFYTHPPIADRIERLRHM 296
 GMI AL+KLD SEPM VDDASAALYI+DP KK GL+ LFYTHPPI++R+ERLR M
 35 Sbjct: 241 GMIRALQKLDNSEPMHRHVDDASAALYISDEKKKGGLQKLFYTHPPISEVERLRKM 297

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4853> which encodes the amino acid
 35 sequence <SEQ ID 4854>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -9.77 Transmembrane 197 - 213 (192 - 223)
 40 INTEGRAL Likelihood = -8.33 Transmembrane 43 - 59 (33 - 61)
 INTEGRAL Likelihood = -3.82 Transmembrane 153 - 169 (153 - 174)

----- Final Results -----

45 bacterial membrane --- Certainty=0.4906(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

50 >GP:AAB70525 GB:AF017421 putative heat shock protein HtpX [Streptococcus gordonii]
 Identities = 208/298 (69%), Positives = 257/298 (85%), Gaps = 1/298 (0%)

Query: 1 MLYQQISQNKQRTVLLVGGFFALLALIGASAGYLLLDNYAMGLVLALVIGVIYATSMIFQ 60
 ML++QI+ NK+RT LLV FFALLALIGA+AGYL +++ G+++A +IG+IYA +MIFQ
 Sbjct: 1 MLFEQIAANKRRTWFLVAVFALLALIGAAAGYLMWNSPLGGVIIAFIIGLIYAITMIFQ 60

55 Query: 61 STSLVMSMNAREVTEKEAPGFFHIVEDMAMVAQIPMPRVFIITEDPSLNAFATGSSPQNA 120
 ST +VMSMN AR+V+E+EAP +HIV+DMAMVAQIPMPRV+I+ED S NAFATGS+P+NA
 Sbjct: 61 STEVVMSMNGARQVSEQEAPELYHIVQDMAMVAQIPMPRVYIVEDDSPNAFATGSPEN 120

60 Query: 121 AVAATTGLLVMNRREELEGVIGHEISHIRNYDIRISTIAVALASAVTVISSIGGRMLWYG 180
 AVAATTGLL +MNRREELEGVIGHE+SHIRNYDIRISTIAVALASA+T+ISS+ GRM+WYG
 Sbjct: 121 AVAATTGLLRLMNRREELEGVIGHEVSHIRNYDIRISTIAVALASAITMISSVAGRMMWYG 180

Query: 181 GGSRRQRDDGDDVLRITLLSLLSLLAPLIVASLIQLAISRQREYLADASSVELTRNP 240
 GG RR+ D DD L ++ L+ SL++++LAPL A+L+QLAISRQRE+LADASSVELTRNP

-1749-

Sbjct: 181 GG-RRRNRDDDSGLGLMLVFSLTATIIAPLAATLVQLAISRQREFLADASSVELTRNP 239

Query: 241 QGMKALEKLQLSQPMKHPVDDASAALYINEPRKKRSFSSSLFSTHPPIEERIERLKNM 298

QGM+AL+KL S+PM VDDASAALYI++P+KK LF THPPI ER+ERL+ M

Sbjct: 240 QGMIRALQKLDNSEPMHRHVDDASAALYISDPKKKGGLQKLFYTHPPTISERVERLRKM 297

An alignment of the GAS and GBS proteins is shown below.

Identities = 233/298 (78%), Positives = 262/298 (87%), Gaps = 2/298 (0%)

10 Query: 1 MLYQQIASNKRTIVLLIVFFCLLAAIGAAGYLVLSYQFGLVLALIIIGVIYAVSMIFQ 60
 MLYQQI+ NK++TVLL+ FF LLA IGA+ GYL+L +Y GLVLAL+IGVIYA SMIFQ
 Sbjct: 1 MLYQQISQNKQRTIVLLVGFALLALIGASAGYLLLDNYAMGLVLALVIGVIYATSMIFQ 60

15 Query: 61 STNVMSMNNAREVTEDEAPNYFHIVEDMAMIAQIPMPRVFIVEDDSLNAFATGSKPEN 120
 ST++VMSMNNAREVTE EAP +FHIVEDMAM+AQIPMPRVFI+ED SLNAFATGS P+NA
 Sbjct: 61 STSLVMSMNNAREVTEKEAPGFFHIVEDMAMVAQIPMPRVFIIEEDPSLNAFATGSSPQNA 120

20 Query: 121 AVAATTGLLAVMNRREELEGVIGHEVSHIRNYDIRISTIAVALASAVTLISSIGSRMLFYG 180
 AVAATTGLL VMNRREELEGVIGHE+SHIRNYDIRISTIAVALASAVT+ISSIG RML+YG
 Sbjct: 121 AVAATTGLLEVMNRREELEGVIGHEISHIRNYDIRISTIAVALASAVTVISSIGSRMLWYG 180

25 Query: 181 GG--RRDDDDREDGGNIIIVLIFSILSLIAPLAASLVQLAISRQREYLADASSVELTRNP 238
 GG R+RDD +D I+ L+ S+LSL+LAPL ASL+QLAISRQREYLADASSVELTRNP
 Sbjct: 181 GGSRRQRDDGDDVLRITITLLLSLLSLLAPLVASLIQLAISRQREYLADASSVELTRNP 240

Query: 239 QGMISALEKLDRESEPMGHVDDASAALYINDPTKKEGLKSLFYTHPPIADRIERLRHM 296
 QGM I ALEKL S+PM HPVDDASAALYIN+P KK SLF THPPI +RIERL++M
 Sbjct: 241 QGMKALEKLQLSQPMKHPVDDASAALYINEPRKKRSFSSSLFSTHPPIEERIERLKNM 298

A related GBS gene <SEQ ID 8847> and protein <SEQ ID 8848> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10

McG: Discrim Score: 9.61

GvH: Signal Score (-7.5): -0.97

Possible site: 25

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 3 value: -11.30 threshold: 0.0

INTEGRAL Likelihood = -11.30 Transmembrane 195 - 211 (190 - 221)

INTEGRAL Likelihood = -11.09 Transmembrane 43 - 59 (31 - 62)

INTEGRAL Likelihood = -3.61 Transmembrane 153 - 169 (153 - 174)

PERIPHERAL Likelihood = 5.89 87

modified ALOM score: 2.76

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5522(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

73.8/88.3% over 296aa

imported

SP|O30795| PUTATIVE HEAT SHOCK PROTEIN HTPX. Insert characterized

GP|2407215|gb|AAB70525.1||AF017421 putative heat shock protein HtpX {Streptococcus gordonii} Insert characterized

PIR|T48855|T48855 probable heat shock protein HtpX - Streptococcus gordonii Insert characterized

ORF02338(301 - 1188 of 1488)

SP|O30795|HTPX_STRGC(1 - 297 of 297) PUTATIVE HEAT SHOCK PROTEIN
 HTPX.GP|2407215|gb|AAB70525.1||AF017421 putative heat shock protein HtpX {Streptococcus gordonii}PIR|T48855|T48855 probable heat shock protein HtpX [imported] - Streptococcus gordonii

-1750-

%Match = 44.0
 %Identity = 73.7 %Similarity = 88.2
 Matches = 219 Mismatches = 34 Conservative Sub.s = 43

```

5      141      171      201      231      261      291      321      351
      NFLFTSVI*HNNIQL*CEIRNFPK*YCWKTIWVQTKPILRNS*RRKRSKSFLL*LLIEKGERLLLYQQIASNKRKTVVLL
                                         :|::|||:|::| |
                                         MLFEQIAANKRRTWFL
                                         10

10     381      411      441      471      501      531      561      591
      IVFFCLLAAIGAAGVYLVLGSYQFGLVLALIIGVVIYAVSMIFQSTNVVMSMNNAREVTEDEAPNYFHIVEDMAMIAQIPM
      : || ||| ||| ||| : | :|::|||:|::| | :|::|||:|::| | :|::|||:|::| |
      VAFFALLALIGAAAGYLWMNSPLGGVIAFIIGLIYAITMIFQSTEVVMSMNGARQVSEQEAPELYHIVQDMAMVAQIPM

15     30      40      50      60      70      80      90

      621      651      681      711      741      771      801      831
      PRVFIVEDDSLNAFATGSKPENAAVAATTGLLAVMNRREELEGVIGHEVSHIRNYDIRISTIAVALASAVTLISSIGSRML
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      PRVYIVEDDSPNAFATGSPENAAVAATTGLRLMNRREELEGVIGHEVSHIRNYDIRISTIAVALASAITMISSVAGRMM

20     110     120     130     140     150     160     170

      861      888      918      948      978      1008     1038     1068
      FYGGRRRRDDREDGG-NILVLIIFSILSLILAPLAASLVQLAISRQREYLADASSVELTRNPQGMISALEKLRSEPMGH
      : ||| ||| : | :|::|||:|::| | :|::|||:|::| | :|::|||:|::| | :|::|||:|::| |
      WYGGRRRRNRDDDSGLGLMLVFSLIAIILAPLAATLVQLAISRQREFLADASSVELTRNPQGMIRALQKLDNSEPMHR

25     190     200     210     220     230     240     250

      1098     1128     1158     1188     1218     1248     1278     1308
      PVDDASAALYINDPTKKEGLKSLFYTHPPIADRIERLRHM*SLTKRRVAMPCVLFF*DKACKT*YNMTYTIKGDGTCYLQ
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      HVDDASAALYISDPKKKGGLQKLFYTHPPISEVERLRKM

30     270     280     290
  
```

SEQ ID 8848 (GBS179) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 175 (lane 11; MW 58kDa).

GBS179-GST was purified as shown in Figure 227, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1571

A DNA sequence (GBSx1665) was identified in *S.agalactiae* <SEQ ID 4855> which encodes the amino acid sequence <SEQ ID 4856>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood =-15.44    Transmembrane    4 - 20 ( 1 - 27)

45     ----- Final Results -----
           bacterial membrane --- Certainty=0.7177(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50
  
```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAG23700 GB:AF017421 LemA-like protein [Streptococcus gordonii]
  Identities = 124/182 (68%), Positives = 152/182 (83%)

55     Query: 1  MGMTILIAIALFVIWLVAYNSLVRSMHTKESWSQIDVQLKRRNDLIPNLIETVKGYA 60
           M +I IA+I + V+++I YNSLVR+RM T+E+WSQIDVQLKRRNDL+PNLIETVKGY
      sbjct: 1  MSFIITIAIVVIVLFFVISVYNSLVRARMQTQEAWSQIDVQLKRRNDLIPNLIETVKGYG 60

      Query: 61  AYEGKTKLEKIAELRAQVAKANTPAEAMTASNELTRQLSSILAVAENYPDLKANNSEFVKLQ 120
  
```


-1751-

YE TLEK+ +LRAQVA A++PA+AM AS+ LTRQ+S I AVAE+YPDLKAN +++KLQ
 Sbjet: 61 KYEQATLEKVTQLRAQVASASSPADAMKASDALTRQISGIFAVAESYPDLKANENYLKLQ 120

Query: 121 EELTNTENKISYSRQLYNTTTSNYNVKLETFPSNIVGKLFQKPSQFLETPEEEKEVPKV 180
 5 EELTNTENKISYSRQLYN+ NYNVKL+ FPSN++ +F F+P+ FL TPEEEK VPKV
 Sbjet: 121 EELTNTENKISYSRQLYNSVAGNYNVKLQAFPSNVIAGMFAFRPADFLSTPEEEKAVPKV 180

Query: 181 SF 182
 F
 10 Sbjet: 181 DF 182

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4857> which encodes the amino acid sequence <SEQ ID 4858>. Analysis of this protein sequence reveals the following:

Possible site: 15
 15 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC44350 GB:U66186 LemA [Listeria monocytogenes]
 25 Identities = 91/181 (50%), Positives = 121/181 (66%), Gaps = 2/181 (1%)

Query: 5 LIILVVLGVLALWLMISYNSLVKSRMHTKEAWSQIDVQLKRRNDLIPNLIETVKGYSYE 64
 +I + V+ +L L YNSLVK R E W+QIDVQLKRR DLIPNL+ETVKGYA +E
 Sbjet: 5 IIAIAVVVLVLIYFGLYNSLVKYNRVRDETWQIDVQLKRRFDLIPNLVETVKGYAKHE 64

30 Query: 65 QKTFEKITDLRARVAN--ASTPQETMAASNELSKQVTSLFAVAENYPDLKANENFLKLQE 122
 ++T ++ + R ++ A Q + A N LS + S+FA+ E YPDLKAN +F++LQ
 Sbjet: 65 KETLTQVIEARNKMEVVPADNRQGEADNMLSGALKSIFALGEAYPDLKANTSFIELQH 124

35 Query: 123 ELTNTENKISYSRQLYNSTTSNYNLQLESFPSNIAGKLFQKPSQFLETPEEEKEVPKVEF 183
 ELT TENK++YSRQLYN+T YN +++S P+NI KL F + L PE E+ PKVEF
 Sbjet: 125 ELTTTENKVAYSRLYNNTVTMTYNKQSVPTNIVAKLHNFTERDMLSIPEVERVAPKVEF 185

An alignment of the GAS and GBS proteins is shown below.

40 Identities = 135/181 (74%), Positives = 165/181 (90%)

Query: 4 MILIAIIALFVIWLIVAYNSLVSRMHTKESWSQIDVQLKRRNDLIPNLIETVKGYAAYE 63
 +I++ ++ + +WL+++YNSLV+SRMHTKE+WSQIDVQLKRRNDLIPNLIETVKGYA+YE
 Sbjet: 5 LIILVVLGVLALWLMISYNSLVKSRMHTKEAWSQIDVQLKRRNDLIPNLIETVKGYSYE 64

45 Query: 64 GKTLEKIAELRAQVAKANTPAEAMTASNELTRQLSSILAVAENYPDLKANNFVKLQEEL 123
 KT EKI +LRA+VA A+TP E M ASNEL++Q++S+ AVAENYPDLKAN +F+KLQEEL
 Sbjet: 65 QKTFEKITDLRARVANASTPQETMAASNELSKQVTSLFAVAENYPDLKANENFLKLQEEL 124

50 Query: 124 TINTENKISYSRQLYNTTTSNYNVKLETFPSNIVGKLFQKPSQFLETPEEEKEVPKVSFDF 184
 TINTENKISYSRQLYN+TTSNYN++LE+FPSNI GKLFQKPS+FL+TPE EKEVPKV F+F
 Sbjet: 125 TINTENKISYSRQLYNSTTSNYNLQLESFPSNIAGKLFQKPSQFLETPEEEKEVPKVEFNF 185

55 A related GBS gene <SEQ ID 8849> and protein <SEQ ID 8850> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0
 McG: Discrim Score: 14.63
 GvH: Signal Score (-7.5): -3.19
 Possible site: 20
 60 >>> Seems to have an uncleavable N-term signal seq

ALOM program count: 1 value: -15.44 threshold: 0.0
 INTEGRAL Likelihood = -15.44 Transmembrane 4 - 20 (1 - 27)

-1752-

PERIPHERAL Likelihood = 8.86 146
modified ALOM score: 3.59

*** Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.7177(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

51.4/68.9% over 183aa

Listeria monocytogenes

EGAD|149857| LemA protein Insert characterized

GP	1519287	gb	AAC44350.1	U66186	LemA Insert characterized
----	---------	----	------------	--------	---------------------------

ORF01545 (301 - 846 of 1152)

EGAD 149857 159923	(2	~	185	of	185)	LemA	protein	{ <i>Listeria monocytogenes</i> }
--------------------	----	---	-----	----	------	------	---------	-----------------------------------

GP|1519287|qb|AAC44350.1|U66186 LemA {*Listeria monocytogenes*}

%Match = 23.8

%Identity = 51.4 %Similarity = 68.9

Matches = 94 Mismatches = 56 Conservative Sub.s = 32

42 72 102 132 162 192 222 252

CFK*TSSLVIAVRLIFSEHSTRSLK*VSNCFCLSVSVIPCSIRT**NAGGVIVNLNFYIV**LYFITNPNNGNRRFL

282 312 342 372 402 432 462 492
I*RKLL*WKKCKGATTMGIMILIAITLFIWIWLIVAYNSLVRSRMTKESWSQIDVQLKRNDLIPNLLETVKGYAAYEG
 : | : | :: : : : ||||| : | :|:::||||| :|||:||||| :
MIGWIIATAVVVLVLVIYFGLYNLSLVKYRNVDETWAQIDVQLKRFDLIPNLVETVKGYAKHEK
 10 20 30 40 50 60

KTLEKIAELRAQVAK--ANTPAEAMTASNELTRQLSSILAVAENYPDLKANNSFVKLQEELTTNTNKISYSRQLYNTTS

:|| :| |:: |: : ||| :| ::||| |:::|| |||:::|||:

ETLTVQVI EARNMMEVPADNRQQGLI EADNMLSGALKSIFALGEAYPDLKANTSFI ELQHETTNTNKVAISRQLYNTVM

80 90 100 110 120 130 140

756 786 816 846 876 906 936 966
NYNVKLETFPSNIVGKLGFGPKSPQFLETPEEKEVPKVSPDF*LRRERGFCCINKLQVIREKQLSC*LSSSVF*QLLEQL
|| ::| :|| | | | | | : ||| |
TYNTKVQSVPTNIVAKLHNFTERDMLSIPEVERVAPKVEF

160 170 180

45 SEQ ID 4856 (GBS42) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 2; MW 21.8kDa) and in Figure 168 (lane 5-7; MW 36kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 8; MW 46kDa). Purified Thio-GBS42-His is shown in Figure 244, lane 11.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
50 vaccines or diagnostics.

Example 1572

A DNA sequence (GBSx1666) was identified in *S.agalactiae* <SEQ ID 4859> which encodes the amino acid sequence <SEQ ID 4860>. This protein is predicted to be glucose inhibited division protein b (gidB). Analysis of this protein sequence reveals the following:

Possible site: 47

```
>>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

-1753-

bacterial cytoplasm --- Certainty=0.2430(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 10079> which encodes amino acid sequence <SEQ ID 10080> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB16137 GB:Z99124 glucose-inhibited division protein [Bacillus subtilis]
 Identities = 130/239 (54%), Positives = 170/239 (70%), Gaps = 4/239 (1%)

10 Query: 5 MTPQAFYQVLIHGITLTDKQKQFETYFRLLVEWNEKINLTAITDKEEVYKHFYDSIA 64
 M + F L E GI+L+ +Q +QFE Y+ +LVEWNEKINLT+IT+K+EVYKHFYDSI
 Sbjct: 1 MNIREFTSGLAEKGISLSPRQLEQFELYDMLVEWNEKINLTSITEKKEVYKHFYDSIT 60

15 Query: 65 PILQGYID-NSPLSILDIGAGAGFPSIPMKILYPEIDITIIDSINKRINFLNILANELEL 123
 Y+D N +I D+GAGAGFPS+P+KI +P + +TI+DSLNRKI FL L+ L+L
 Sbjct: 61 AAF--YVDFNQVNTICDVGAGAGFPSLPIKICFPHLHVTIVDSLNRKITFLEKLSALQL 118

20 Query: 124 SGVHFFHGRAEDFGQDRVFRAKFDIVTARAVAKMQVLAEITPFLKVNRLIALKAAAAE 183
 F H RAE FGQ + R +DIVTARAVA++ VL+EL +P +K NG +ALKAA+AE
 Sbjct: 119 ENTTFCHDRAETFGQRKDVRESYDIVTARAVARLSVLSELCLPLVKKNGLFVALKAASAE 178

25 Query: 184 EELISAEKALKITLFSQVTVNKNYKLP-NGDDRNITIVSKKKETPNKYPRKAGTPNKKPL 241
 EEL + +KA+ TL ++ ++KLP DRNI ++ K K TP KYPRK GTPNK P+
 Sbjct: 179 EELNAGKKAITTLGGELENIHSFKLPIEESDRNIMVIRIKINTPKKYPRKPGTPNKSPI 237

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4861> which encodes the amino acid sequence <SEQ ID 4862>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4862(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 170/237 (71%), Positives = 202/237 (84%)

40 Query: 5 MTPQAFYQVLIHGITLTDKQKQFETYFRLLVEWNEKINLTAITDKEEVYKHFYDSIA 64
 MTPQ FY+ L E G +L+ KQK+QF+TYF+ LVEWN KINLTAIT++ EVYKHFYDSIA
 Sbjct: 1 MTPQDFYRTLEEDGFSLSSKQKEQFDITYFKSLVEWNTKINLTAITEENEVYKHFYDSIA 60

45 Query: 65 PILQGYIDNSPLSILDIGAGAGFPSIPMKILYPEIDITIIDSINKRINFLNILANELELS 124
 PILQG++ N P+ +LDIGAGAGFPS+PMKIL+P +++TIIDSINKRI+FL +LA EL L
 Sbjct: 61 PILQGFLANEPIKLLDIGAGAGFPSLPMKILFPNLEVTTIIDSINKRISFLTLLAQELGLE 120

50 Query: 125 GVHFFHGRAEDFGQDRVFRAKFDIVTARAVAKMQVLAEITPFLKVNRLIALKAAAAEE 184
 VHFFHGRAEDFGQD+ FR +FD+VTARAVA+MQVL+ELTIPFLK+ G+LIALKA AA++
 Sbjct: 121 NVHFFHGRAEDFGQDKAFRGQFDVVTARAVARMQVLSELTIPFLKIGGKLIALKAQAADQ 180

55 Query: 185 ELISAEKALKITLFSQVTVNKNYKLPNGDDRNITIVSKKKETPNKYPRKAGTPNKKPL 241
 EL A+ AL LF +V N +Y+LPNGD R ITIV KKKETPNKYPRKAG PNNKKPL
 Sbjct: 181 ELEEAKNALCLLFGKVIKNHSYQLPNGDSRFITIVEKKKETPNKYPRKAGLPNNKKPL 237

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1754-

Example 1573

A DNA sequence (GBSx1667) was identified in *Sagalactiae* <SEQ ID 4863> which encodes the amino acid sequence <SEQ ID 4864>. Analysis of this protein sequence reveals the following:

Possible site: 13
 5 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 10 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 15 vaccines or diagnostics.

Example 1574

A DNA sequence (GBSx1668) was identified in *Sagalactiae* <SEQ ID 4865> which encodes the amino acid sequence <SEQ ID 4866>. This protein is predicted to be v-type sodium ATP synthase subunit j. Analysis of this protein sequence reveals the following:

Possible site: 45
 20 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -10.14	Transmembrane	371 - 387 (362 - 391)
INTEGRAL	Likelihood = -7.48	Transmembrane	200 - 216 (190 - 217)
INTEGRAL	Likelihood = -4.94	Transmembrane	425 - 441 (423 - 446)
25 INTEGRAL	Likelihood = -4.67	Transmembrane	327 - 343 (325 - 349)
INTEGRAL	Likelihood = -3.77	Transmembrane	81 - 97 (81 - 98)
INTEGRAL	Likelihood = -2.66	Transmembrane	140 - 156 (139 - 157)
INTEGRAL	Likelihood = -1.33	Transmembrane	55 - 71 (53 - 71)
INTEGRAL	Likelihood = -0.27	Transmembrane	247 - 263 (247 - 263)
30 INTEGRAL	Likelihood = -0.11	Transmembrane	165 - 181 (165 - 181)

----- Final Results -----
 bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10081> which encodes amino acid sequence <SEQ ID 10082> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:BAA04279 GB:D17462 Na+ -ATPase subunit J [Enterococcus hirae]
 Identities = 170/461 (36%), Positives = 262/461 (55%), Gaps = 28/461 (6%)

Query: 12 KTMVARKLSISFIIVILLGSILLSLPFQYANAPKTHYIDHLFTTVSMVCVTGLSVFPPI 71
 K +S + ++ F +IL G LL+LP F + TH+ID LFT S VCVTGL+

45 Sbjct: 10 KRLSPVQLIAAGFFILILFGGSLTLPPFS-RSGESTHFIDALFTATSACVVTGLTTLNT 68

Query: 72 SKVYNGWGQIVAILLMQTGGLGLVTLMSLSYYTLRRKMSLNDQTLLQSAITYNSSTDLLK 131
 ++ +N GQ + + L++ GGLG + + L + ++K+S + + +L+ A+ + + K

50 Sbjct: 69 AEHWSAGQFLIMTLIEIGGLGFMPIPLFFAIKKKISFSMRIVLKEALNLEMSGVIK 128

Query: 132 YLYMIFKVTLTLEVLAAISILAIIDFIPRFLGHGIFNSIFLAVSAFCNAGFDNLEATSLAQ 191
 + I K + ++V+ A L++ FIP FG GI+ SIF AVS+FCNAGFD L + LA

Sbjct: 129 LMIYILKFAVVIQVIGAVALSVMFIPFGWAKGIWFSIFHAVSSFCNAGFDLLGDSLLAD 188

-1755-

Query: 192 FKLNLPLVNIIVCFLLIISGGLGFAVWKDLIEATTQTSKGPGLIKTFPKRLSNHSLVLKT 251
 + N + ++V LII+GGLGF VW+D++ + H+ K+++ HSK+ L
 Sbjct: 189 -QTNVYLIMVVSALIIAGGLGFIVWRDIL-----SYHR-----VKKITLHSLKVALSV 234

5 Query: 252 TTIILLTGTLISWLLLEFGNFRTIANLSLPKQLMVSFFQTVMTAGFSTIDYTQTDFAFN 311
 T ++L+ G +L +L+ N T+ + ++L +FF +VT RTAG+ +IDY Q A
 Sbjct: 235 TALLLIGGFIL-FLITERNGLTLVKGTFTERLANTFFMSVTPRTAGYYSIDYLQMSHAGL 293

10 Query: 312 LVYIIQMLIGGAPGGTAGGFKVTVIAILLLLFKAELSGQSQVTFHYRTIPSSIIKQTLSE 371
 ++ + M IGG G TAGG K T + ILL+ A G+++ RTI + + L
 Sbjct: 294 ILTMFLMYIGGTSGSTAGGLKTTTLGILLIQMHAMFKGKTRAEAFGRTRQAAV---LRA 350

15 Query: 372 LTTTTFII--LISGYLLLELNPIDPFS----LFFEASSALATVGVTMNTTNQLTLGGRI 425
 LT FF+ L +++L + I S + FE SA TVG+TM T LTL G++
 Sbjct: 351 LTLFFVTLSLCVVAIMVLSVTETIPKTSGLIEYIAFEVFSAFGTVGLTMGLTPDLTLIGKL 410

20 Query: 426 VIMFLMFIGRVGPITVLLSILQK---KEKEIHYAETEIIIG 463
 VI+ LM+IGRVG +TV+LS+L K E Y E I+LG
 Sbjct: 411 VIISLMYIGRVGIMTVVLSLLVKANRAEANYKYPEESIMLG 451

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4867> which encodes the amino acid sequence <SEQ ID 4868>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -15.12 Transmembrane 371 - 387 (364 - 396)
 INTEGRAL Likelihood = -7.32 Transmembrane 20 - 36 (18 - 42)
 INTEGRAL Likelihood = -6.53 Transmembrane 425 - 441 (417 - 446)
 INTEGRAL Likelihood = -6.16 Transmembrane 89 - 105 (81 - 106)
 30 INTEGRAL Likelihood = -5.79 Transmembrane 200 - 216 (196 - 223)
 INTEGRAL Likelihood = -3.35 Transmembrane 140 - 156 (139 - 157)
 INTEGRAL Likelihood = -3.03 Transmembrane 55 - 71 (53 - 74)
 INTEGRAL Likelihood = -3.03 Transmembrane 247 - 263 (246 - 264)
 INTEGRAL Likelihood = -1.12 Transmembrane 393 - 409 (393 - 409)
 35 INTEGRAL Likelihood = -0.11 Transmembrane 165 - 181 (165 - 181)

----- Final Results -----
 bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAA04279 GB:D17462 Na+ -ATPase subunit J [Enterococcus hirae]
 Identities = 168/466 (36%), Positives = 260/466 (55%), Gaps = 26/466 (5%)

45 Query: 6 MKRSFIKSLSVTQRLTFSSFAIVILIGTLLLSMPFTHYQNGPNTVYLDHFFENVVSMVCVTG 65
 MK+ K LS Q + F I+IL G LL++PF ++G +T ++D F S VCVTG
 Sbjct: 4 MKKRVKRLSPVQLIAAGFFILILFGGSLTLPLFFS-RSGESTHFIDALFTATSAVCVTG 62

50 Query: 66 LSVVPVAEYVNGIGQTIAMALMQIGCLGLVTLIAVSTFAL-KRKMRLSDQTLQLQSALNRG 124
 L+ + AE +N GQ + M L++IG LG + +I + FA+ K+K+ S + +L+ ALN
 Sbjct: 63 LTTLNTAEHWNSAGQFLIMTLIEIGGLGFM-MIPILFFAIKKKISFSMRIVLKEALNLE 121

55 Query: 125 DSKDLKHYLFFAYKVTFSLFAFAIVIMIDFIPRFGWKNGIFNSIFLAVSAFCNAGFDNL 184
 + + + + K ++ A+ + + FIP FGW GI+ SIF AVS+FCNAGFD L
 Sbjct: 122 EMSGVIKLMIYILKFAVVIQVIGAVALSUVFIPEFGWAKGIWFSIFHAVSSFCNAGFDLL 181

60 Query: 185 GSSSLKDFMLNPTLNVIITFLIISGGLGFAVWDLGVAFKKYFFERPHCYGATFRKLSNQ 244
 G S L D N L ++++ LII+GGLGF VW D+ +++ + +K++
 Sbjct: 182 GDSLLAD-QTNVYLIMVVSALIIAGGLGFIVWRDI-LSYHR-----VKKITLH 227

65 Query: 245 SRLVLQTTAVILFLGTFLTWFLEKDNSTIANFSLHQQLMVSFFQTVMTAGFATISYN 304
 S++ L TA++L +G F+ + + + N T+ + ++L +FF +VT RTAG+ +I Y
 Sbjct: 228 SKVALSVTALLL-IGGFILFLITERNGLTLVKGTFTERLANTFFMSVTPRTAGYYSIDYL 286

Query: 305 DTLAPTNIILYMIQMVIGGAPGGTAGGIKVTAAITFLLFKAELSGQSEVTFNRNRIANKT 364

-1756-

IL M M IGG G TAGG+K TT I + A G++ R I
 Sbjct: 287 QMSHAGLILTMFLMYIGGTSGSTAGGLKTTTLGILLIQMHAMFKGKTRAEAFGRITIRQAA 346

Query: 365 IKQTMVTLIFFFAVLMIIGFILLLSVEPHIAPIP---LLFESISAIATVGVSMDLTPQLS 420
 + + +T L F L + I++LSV I + FE SA TVG++M LTP L+

Sbjct: 347 VLRLALT-LFFVTLSLCVVIMVLSVTETIPKTSIEYIAFEVFSAFGTVGLTMGLTPDLT 405

Query: 421 TAGRLIVIVLMFVGRVGPITVLISLI---QRKEKTIQYATTDILVG 463
 G+L++I LM++GRVG +TV++SL+ R E +Y I++G

Sbjct: 406 LIGKLVIIISLMYIGRVGIMTVVLSLLVKANRAEANYKYPEESIMLG 451

An alignment of the GAS and GBS proteins is shown below.

Identities = 275/462 (59%), Positives = 351/462 (75%), Gaps = 1/462 (0%)

Query: 2 GASMKHFFDYKTMSVARKLSISFIAVILLGSIILSLPIFOYANAPKTHYIDHLFTTVSMV 61
 G +MK F K++SV ++L+ SF VIL+G++LLS+P Y N P T Y+DH F VSMV

Sbjct: 3 GGNMKRSF-IKSLSVTQRLTFSFAIVILIGTLLSMPFTHYQNGPNTVYLDHFFNVVSMV 61

Query: 62 CVTGLSVFPISKVYNGWQIVAILMQTGGGLGLVTLMSLSYTLRRKMSLNDQTLLQSAI 121
 CVTGLSV P+++VYNG GQ +A+ LMQ G LGLVTL+++S + L+RKM L+DQTLLQSA+

Sbjct: 62 CVTGLSVVPVAEVYNGIGQTIAMALMQIGCLGLVTLIAVSTFALKRKMRLSDQTLLQSAL 121

Query: 122 TYNSTDLKKYLYMIFKVTLTLEVLAAASILAIDFIPRFGLGHGIFNSIFLAVSAFCNAGF 181
 S DLK YL+ +KVT +LE AA ++ IDFIPRFG +GIFNSIFLAVSAFCNAGF

Sbjct: 122 NRGDSKDLKHLYFFAYKVTFSLFAAIVIMIDFIPRFGWKNGIFNSIFLAVSAFCNAGF 181

Query: 182 DNLEATSLAQFKLNPLVNIIVCFLLIISGGLGFAVWKDLIEATIQTSHKGPKLKTFPKRL 241
 DNL ++SL F LNP +N+I+ FLIISGGLGFAVW DL A + + P ++L

Sbjct: 182 DNLGSSSLKDFMLNPTLNVIITFLIISGGLGFAVWDLGVAFKKYFFERPHCYGATFRKL 241

Query: 242 SNHSLVLKTTTIIILLTGTLSSWLEFGNFRTIANLSLPKQLMVSPFQTVIMRTAGFSTI 301
 SN S+LVL+TT +IL GT L+W LE N +TIAN SL +QLMVSPFQTVIMRTAGF+TI

Sbjct: 242 SNQSRVLVQTAVILFLGTFLTWFLKDNSTIANFSLHQQLMVSPFQTVIMRTAGFATI 301

Query: 302 DYTQTFATNLVYIIQMLIGGAPGGTAGGFKVTVAIILLLLFKAELSGQSQVTFHYRTIP 361
 Y T TN++Y+IQM+IGGAPGGTAGG KVT AI LLFKAELSGQS+VTF R I

Sbjct: 302 SYNDTLAPTNLILYMIQMVIGGAPGGTAGGKVTAAITFLLFKAELSGQSEVTFNRRIIA 361

Query: 362 SSIKQTLISILTFFFIILISGYLLLELNPIDPFLSFFEASSALATVGVIMNTTNQLTL 421
 + IKQT+++L FFF +L+ G++LLL + PHI P L FE+ SA+ATVGV+M+ T QL+

Sbjct: 362 NKTIKQTMVLIFFFAVLMIIGFILLLSVEPHIAPILLFESISAIATVGVSMDLTPQLST 421

Query: 422 GGRIVIMFLMFIGRVGPITVLLSILQKKEKEIHYAETEIIIG 463
 GR++++ LMF+GRVGPITVL+S++Q+KEK I YA T+I++G

Sbjct: 422 AGRLIVIVLMFVGRVGPITVLISLIQRKEKTIQYATTDILVG 463

A related GBS gene <SEQ ID 8851> and protein <SEQ ID 8852> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
 McG: Discrim Score: 0.86
 GvH: Signal Score (-7.5): 0.64
 Possible site: 45
 >>> Seems to have a cleavable N-term signal seq.

ALOM program count: 9 value: -10.14 threshold: 0.0

INTEGRAL Likelihood = -10.14 Transmembrane 371 - 387 (362 - 391)
 INTEGRAL Likelihood = -7.48 Transmembrane 200 - 216 (190 - 217)
 INTEGRAL Likelihood = -4.94 Transmembrane 425 - 441 (423 - 446)
 INTEGRAL Likelihood = -4.67 Transmembrane 327 - 343 (325 - 349)
 INTEGRAL Likelihood = -3.77 Transmembrane 81 - 97 (81 - 98)
 INTEGRAL Likelihood = -2.66 Transmembrane 140 - 156 (139 - 157)
 INTEGRAL Likelihood = -1.33 Transmembrane 55 - 71 (53 - 71)
 INTEGRAL Likelihood = -0.27 Transmembrane 247 - 263 (247 - 263)
 INTEGRAL Likelihood = -0.11 Transmembrane 165 - 181 (165 - 181)
 PERIPHERAL Likelihood = 2.49 308

modified ALOM score: 2.53

*** Reasoning Step: 3

----- Final Results -----

```
5      bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1758-

Example 1575

A DNA sequence (GBSx1669) was identified in *S.galactiae* <SEQ ID 4869> which encodes the amino acid sequence <SEQ ID 4870>. This protein is predicted to be TrkA (ktrA). Analysis of this protein sequence reveals the following:

5 Possible site: 19
 >>> Seems to have a cleavable N-term signal seq.

 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC46144 GB:AF001974 putative TrkA [Thermoanaerobacter
 ethanolicus]
 Identities = 69/177 (38%), Positives = 110/177 (61%), Gaps = 2/177 (1%)

Query: 8 VLGLGIFGQTLAQELSNFEQDVIAIDSNPEN--VQVAEVVTKAAIGDITDLAFLKHIGI 65
 V+GLG FG +LA+ L DV+ ID + E VQA+ +VT A D TD LK + +
 20 Sbjct: 6 VIGLGSFGISLAKTLYEMGNDVLVIDEDEEEELVQAMNGLVTHAVRADATDENVLKSLRV 65

Query: 66 SDCDTVIIATGNSLESSVLAVMHCKKLGVPQVIKARNLVYEEVLYEIGADLVISPERES 125
 + D I+A G ++ESS++ M K+LGV VIAKA N ++ VLY++GAD V+ PE++
 25 Sbjct: 66 KNFDVAIVAIGKNMESSIMVTMLVKELGVKYVIAKAHNEHARVLYKVGADRVVMPEKDM 125

Query: 126 GQNVANLMRNKITDVFQIESDISVIEFKIPKSWVGKTVEQLNIRHKFDNLIGIRK 182
 G VA N+ + + D+ + + S+ E + W GKT++++N+R K+ LN++ ++K
 30 Sbjct: 126 GIRVARNVFSSNLIDLIEFSKEYSIAEILPIEEWFGKTLKEINVREKYGLNVVAVKK 182

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4715> which encodes the amino acid sequence <SEQ ID 4716>. Analysis of this protein sequence reveals the following:

 Possible site: 20
 >>> Seems to have an uncleavable N-term signal seq

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

 Identities = 132/221 (59%), Positives = 176/221 (78%)

Query: 1 MKTKIIGVLGLGIFGQTLAQELSNFEQDVIAIDSNPENVQVAEVVTKAAIGDITDLAFL 60
 +K K +GVLGLGIFG+T+A+ELSNF+QDVIAID +V+ VA++VTKAA+GDITD FL
 45 Sbjct: 2 LKRKTVGVLGLGIFGRTVARELSNFDQDVIAIDIRESHVKEVADLVTKAAVGDITDKEFL 61

Query: 61 KHIGISDCDTVIIATGNSLESSVLAVMHCKKLGVPQVIKARNLVYEEVLYEIGADLVIS 120
 +GI CDTV+IA+GN+LESSVLAVMHCKKLGVP +IAKA+N ++EEVLY IGA VI+
 50 Sbjct: 62 LAVGIEHCDTVVIAAGNNLESSVLAVMHCKKLGVPITIIAKAKNKIFEEVLYGIGATKVIT 121

Query: 121 PERESGQNVANLMRNKITDVFQIESDISVIEFKIPKSWVGKTVEQLNIRHKFDNLIGI 180
 PER+SG+ VA+NL+R I + +E IS+IEF IPKSW G+++ +L++R K++LN+IG+
 55 Sbjct: 122 PERDSGKRVASNLLRRHIESIIYLEHGISMIEFVIPKSWEGQSLSELDVRRKYELNVIGM 181

Query: 181 RKAKNKPVDTEVPINSPLEEGIIILVAIANSDAFQRYDYLGY 221
 R+ + K +DT V PLE I+VAIAN F+++DYLGY
 Sbjct: 182 RQKEVKTLDITNVKPFEPLEPNTIIVAIAANDHTFEKFDYLGY 222

A related GBS gene <SEQ ID 8853> and protein <SEQ ID 8854> were also identified. Analysis of this protein sequence reveals the following:

```

5      Lipop: Possible site: -1      Crend: 3
      McG: Discrim Score:      5.14
      GvH: Signal Score (-7.5): -0.860001
          Possible site: 19
      >>> Seems to have a cleavable N-term signal seq.
      ALOM program   count: 0 value:   1.06 threshold:  0.0
          PERIPHERAL Likelihood =   1.06      192
      modified ALOM score:  -0.71

      *** Reasoning Step: 3

      ----- Final Results -----
15      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear)

```

38.0/61.6% over 182aa

Thermoanaerobacter

ethanolicus

GP|2581796| putative TrkA Insert characterized

ORF02030(322 - 864 of 1269)

GP|2581796|gb|AAC46144.1||AF001974(6 - 188 of 195) putative TrkA {Thermoanaerobacter

ethanolicus}

%Match = 15.5

%Identity = 37.9 %Similarity = 61.5

Matches = 69 Mismatches = 69 Conservative Sub.s = 43

60 90 120 150 180 210 240 270

LISGYLILLELNPIDPFSLFFREASSALATVGVTMTTNQLTLGGRIVIMFLMFIGRVGPITVLLSILQKKEKEIHYAET

35 300 330 360 390 444 474 504

EIILG*KRSFMKTKIIGVLGLGIFGOTLAQELSNEFQDVIAIDSNPEN--VQVAEAVVTKAAIGDITDLAFLKHIGISDC

| :||| || :||: | ||: ||: | ||: ||: | ||: ||: |

MKQFVVIGLGSFGISLAKTYIEMGNDVLVIDEDEEEELVQAMNGLVTHAVRADATDENVLKSLRVKNF

10 20 30 40 50 60

534 564 594 624 654 684 714 744

DTVIIATGNSLESSVLAVMHCKKLGVPQVIAKARNLVXEEVLYEIGADLVISPERESGQNVAANLMRNKITDVFQIESDI

| |:| | :|||: | |:| | |||| | : |||:| | |:|: | || |: : : |: : :

DVAIVAIGNMESSIMVIMLVKELGVKYVIAKAHNELHARVLYKVGADRVMPEKDMGIRVARNVFSSNLIDLIEFSKEY

80 90 100 110 120 130 140

774 804 834 864 894 924 954 984

SVIEFKIPKPSWVGKTVEQLNIRHKFDNLNIGIRKAKNKPVDTEVPINSPLEEXIILVAIANSDAFQRYDYLRIFY*RK*K

|: | : | |||:|:| |:|:|:| : : :

SIAEILPIEEWFGKTLKEINVREKYGLNVVAVKKFNDEIIVSPGAGL

160 170 180 190

SEQ ID 8854 (GBS57) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 6; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 11; MW 51.1kDa) and in Figure 183 (lane 9 & 10; MW 51kDa).

The GBS57-GST fusion product was purified (Figure 99A; see also Figure 195, lane 8) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 99B), FACS (Figure 99C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

-1760-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1576

A DNA sequence (GBSx1670) was identified in *S.agalactiae* <SEQ ID 4871> which encodes the amino acid sequence <SEQ ID 4872>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

10	INTEGRAL	Likelihood = -11.62	Transmembrane	73 - 89 (68 - 96)
	INTEGRAL	Likelihood = -11.30	Transmembrane	254 - 270 (248 - 274)
	INTEGRAL	Likelihood = -4.73	Transmembrane	127 - 143 (124 - 144)
	INTEGRAL	Likelihood = -4.19	Transmembrane	50 - 66 (47 - 67)
	INTEGRAL	Likelihood = -3.29	Transmembrane	25 - 41 (25 - 45)

----- Final Results -----

15	bacterial membrane	---	Certainty=0.5649(Affirmative)	< succ>
	bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
	bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 8855> which encodes amino acid sequence <SEQ ID 8856> was also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 9

McG: Discrim Score: -10.49

GvH: Signal Score (-7.5): -1.14

Possible site: 40

>>> Seems to have no N-terminal signal sequence

ALOM program count: 5 value: -11.62 threshold: 0.0

25	INTEGRAL	Likelihood = -11.62	Transmembrane	73 - 89 (68 - 96)
	INTEGRAL	Likelihood = -11.30	Transmembrane	254 - 270 (248 - 274)
	INTEGRAL	Likelihood = -4.73	Transmembrane	127 - 143 (124 - 144)
30	INTEGRAL	Likelihood = -4.19	Transmembrane	50 - 66 (47 - 67)
	PERIPHERAL	Likelihood = 3.76	201	

modified ALOM score: 2.82

*** Reasoning Step: 3

----- Final Results -----

35	bacterial membrane	---	Certainty=0.5649(Affirmative)	< succ>
	bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
40	bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13178 GB:Z99110 ykoC [Bacillus subtilis]

Identities = 61/226 (26%), Positives = 108/226 (46%), Gaps = 12/226 (5%)

45	Query: 49	FLIVVSLGSLVLFRLAKIKWQQVSFVMTLVVVFAVLNIIMVYLFAPHYGDKIYGSSSLLL 108
		F I++ G L+ + KW + + F +L V+ A K+ + L

Sbjct: 36 FYIIIVAGVLLAAGIPLKKW-----LLFTIPFLILAFGCVWTAADF--GKVFPTPDNLF 87

50	Query: 109	KGIGPYDVTSQELFYLFNLILKYFCTVPLALLFLMTTNPSQFASSL-NQLGLSYKIAYAV 167
		GP + S + +L + C L+++F+ TT+P F SL Q LS K+AY V

Sbjct: 88 FQAGPISINSNDNVSVGISLGRILCFSAISMVFVFTTDPILFMLSIVQOCRLSPKLAYGV 147

55	Query: 168	SLTLRYIPDVQEEFYTTIRRAQEARGIELSKSNLVARIKGNLQIVTPLIFSSLERIDTVA 227
		R++P +++E I++A + RG + +S ++ +I + PL+ S++ + + A

Sbjct: 148 IAGFRFLPLLKDEVQLIQAHKIRGG--AAESGIINKISALKRYTIPLLASAIRKAERTA 205

Query: 228 TAMELRRFGKNKRRTWYSKQSLEKSDIVLIILALASLFVSLYLIHL 273

AME + F ++ RT+Y S+ + D V L L LF +L+ L

Sbjct: 206 LAMESKGFTGSRNRTYYRTLSVNRDWFVFCVLVLL-LFAGSPLVSL 250

60

-1761-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1577

A DNA sequence (GBSx1671) was identified in *S.agalactiae* <SEQ ID 4873> which encodes the amino acid sequence <SEQ ID 4874>. This protein is predicted to be cobalt ABC transporter, ATP-binding protein (cbiO). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.91    Transmembrane  436 - 452 ( 435 - 452)

----- Final Results -----
          bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13179 GB:Z99110 similar to cation ABC transporter
(ATP-binding protein) [Bacillus subtilis]
Identities = 151/483 (31%), Positives = 248/483 (51%), Gaps = 19/483 (3%)

Query: 8  KDFTFYDYDVQSEPTLKGINLSIPKGEKVLILGPSGSGKSTLGHCLNGIIPNTHKGQYSGI 67
      + +F Y+ +P + I+ + KGE VL+LGPSG GKS+L CLNG+ P G SG
Sbjct: 11 BQLSFSYEEDEKPVFQDISFELQKGECVLLLGPSGCGKSSLALCLNGLYPEACDGIQSGH 70

Query: 68 FTINHKNADFDSIYDK-SHLVSTVLQDDPGQFGLTVAEDIAFALENDVVAQEEMASIVE 126
      + K D + + V QDDP QF LTV ++IAF LEN + +EEM +
Sbjct: 71 VFLFKPVTDAETSETITQHAGVVFQDDPGQFCMLTVEDEIAFGLENLQIPKEEMTEKIN 130

Query: 127 MWAKRLIEIAPLLSKRPQDLGGQKQRVSLAGVLVDDSPILLFDEPLANLDPQSGQDIMAL 186
      +L I L K LSGGQKQ+V+LA +L + +++ DEP + LDP S ++ + L
Sbjct: 131 AVLGLRLITHLKEKMISTLSGGQKQKVALACILAMEPELIILDEPTSLDPFSAREFVHL 190

Query: 187 VDRIHQEQDATTIIIEHRLD--VFYERVDVRVLFSDGQIINYNGEPDQLL--KTNFLSEY 242
      + + +E+ + ++IEH+L++ + ER +VL G+ +G L + L +
Sbjct: 191 MKDLQREKGFSLLVIEHQLDEWAPWIERT--IVLDKSGKKALDGLTKNLFQHEAETLKKL 248

Query: 243 GIREPLYISALKNLGYDFEKQNTMTSIDDFDFSELLIPKMRALDLDKHTDKLLSVQHLSV 302
      GI P + L F M + + K +A + +L V LS
Sbjct: 249 GIAIPKVCHLQEKLSMPFTLSKEMLFKEPIAGH--VKKKKA---PSGESVLEVSSLSF 302

Query: 303 SYDLENNTLDDVSFDLYKGQRLAIVGKNGAGKSTLAKALCOFI-PNNATLIYNNEDVSQD 361
      + + D+SF L +G A+VG NG GKSTL L + P + ++ ++ + +
Sbjct: 303 ARG-QQAIKDISFSLREGSLTALVGPNGTGKSTLLSVLASLMKPQSGKILLYDQPLQKY 361

Query: 362 SIKERAERIGYVLQNPQMISQAMVFDEVALGLRLRGFSNDIESRVYDILKVCGLYQFR 421
      KE +R+G+V QNP V+DE+ G + ++ + E + +L+ GL
Sbjct: 362 KEKELRKRMGFVFQNPHEHQFVTDTVYDELLFGQK---ANAETEKKAQHLLQRFGLAHLA 417

Query: 422 NWPISALSFGQKKRVTIASILILNPEVIILDEPTAGQDMKHYTEMMSFLDKLSCDGHITIV 481
      + A+S GQK+R+++A++L+ + +V++LDEPT GQD + E M + ++ +G ++
Sbjct: 418 DHHFPAISQGQKRRLSVATMLMHDVKVLLLEPTFGQDARTAAECMEMIQRIKAEGTAVL 477

Query: 482 MIT 484
      MIT
Sbjct: 478 MIT 480
```

There is also homology to SEQ ID 4416.

SEQ ID 4874 (GBS424d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 2 & 4; MW 77kDa) and in Figure 239 (lane 10; MW 77kDa). It

-1762-

was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 5 & 7; MW 52kDa) and in Figure 182 (lane 4; MW 52kDa). Purified GBS424d-His is shown in Figure 241, lanes 6 & 7. Purified GBS424d-GST is shown in Figure 246, lane 12.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1578

A DNA sequence (GBSx1672) was identified in *S.agalactiae* <SEQ ID 4875> which encodes the amino acid sequence <SEQ ID 4876>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence
  INTEGRAL    Likelihood = -8.12    Transmembrane  39 - 55 ( 35 - 63)
  INTEGRAL    Likelihood = -3.98    Transmembrane  72 - 88 ( 71 - 90)
  INTEGRAL    Likelihood = -3.66    Transmembrane 108 - 124 ( 106 - 127)
  INTEGRAL    Likelihood = -2.34    Transmembrane 182 - 198 ( 181 - 198)
  INTEGRAL    Likelihood = -1.44    Transmembrane 141 - 157 ( 139 - 158)

----- Final Results -----
      bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB59830 GB:AJ012388 hypothetical protein [Lactococcus lactis]
Identities = 109/182 (59%), Positives = 141/182 (76%)

Query: 31  MNNTIKKVVATGIGALFIIIGMLVNIPTPIPTNIQLQYAVLALFAVIYGPVGFFTG 90
          M N++K VVATGIGALF+IIG L+NIPTPIPT+IQLQYAVLALF+ ++GP GF G
Sbjct: 1   MKNNSVKIVVATGIGALFVIIGWLINIPTPIPTNSIQLQYAVLALFSAFGPLAGFLIG 60

Query: 91  FIGHALKDSIQYGSPWWTVLVSGLLGLMIGFFAKKLAIQLSGMTKKDLLLFNVVQVIAN 150
          FIGHALKDS YG+PWWTVWL SGL+GL +GF K+ ++          K+++ FN+VQ +AN
Sbjct: 61  FIGHALKDSFLYGAPWWTVLVGSGLMGLFLGFGVKRESLTQGIFGNKEIIRFNIVQFLAN 120

Query: 151 LIGWSVVAPYGDIFFYSEPASKVFAQGFLSSLVNSITIGVGGTLLLLLAYAKSRPQKGSLS 210
          ++ W ++AP GDI YSEPA+KVF QG ++ LVN++TI V GTLLL YA +R + G+L
Sbjct: 121 VVVWGLIAPIGDILVYSEPANKVFTQGVVAGLVNALTIAVAGTLLLLKYAATRTKSGTLD 180

Query: 211 KD 212
          K+
Sbjct: 181 KE 182

```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8857> and protein <SEQ ID 8858> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 6
McG: Discrim Score:      -5.01
GvH: Signal Score (-7.5): -5.9
Possible site: 50
>>> Seems to have no N-terminal signal sequence
ALOM program count: 5 value: -8.12 threshold: 0.0
  INTEGRAL    Likelihood = -8.12    Transmembrane  31 - 47 ( 27 - 55)
  INTEGRAL    Likelihood = -3.98    Transmembrane  64 - 80 ( 63 - 82)
  INTEGRAL    Likelihood = -3.66    Transmembrane 100 - 116 ( 98 - 119)
  INTEGRAL    Likelihood = -2.34    Transmembrane 174 - 190 ( 173 - 190)
  INTEGRAL    Likelihood = -1.44    Transmembrane 133 - 149 ( 131 - 150)
  PERIPHERAL  Likelihood = 5.78      9
modified ALOM score: 2.12

```

-1763-

*** Reasoning Step: 3

----- Final Results -----

5 bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 ORF02330(367 - 912 of 1212)
 GP|6165407|emb|CAB59830.1||AJ012388(1 - 182 of 182) hypothetical protein {Lactococcus
 lactis}
 %Match = 28.1
 %Identity = 59.9 %Similarity = 78.6
 15 Matches = 109 Mismatches = 39 Conservative Sub.s = 34

102	132	162	192	222	252	282	312
MQVVGVGFIQSVIQDSCETALNSSTDVLTAVAESVFGKK*TNEGLRYSI*DLFWYLILFSIVFQFFLSIRFQISLKYD							
342	372	402	432	462	492	522	552
KIEQIVSDCLSLFFREVFMMNINTIKKVATGIGAAALFIIIGMLVNIPTPIPTNTNIQLQYAVLALFAVIYGPVGFFFTGFI							
: : : : :							
MKNSVKIVVATGIGAAALFVIIGWLINIPTPIPTNTSIQLQYAVLALFSALEGLAGFLGFI							
10 20 30 40 50 60							
582	612	642	672	702	732	762	792
GHALKDSIQYGPWWTWVLVSGLLGLMIGFFAKKLAIQLSGMTKKDLLFNVVQVIANLIGWSVVPYGDIFFYSEPAK							
: : : : : : : : : :							
GHALKDSFLYGAPWWTWVLGSGLMGLFLGFGVKRESLTQGFIGNKEIIRFNIVQFLANVVWGLIAPIGDILVYSEPAK							
80 90 100 110 120 130 140							
822	852	882	912	942	972	1002	1032
VFAQGFLSSLVNSITIGVGGTLLLLAYAKSRPQKGSLSKD*DKRVIYERFY*MEGFYLSI*RSI*TNFKRD*LKHS*R*K							
: : : : : : :							
VFTQGVVAGLVNALTITAVAGTLLKLKLYAATRTKSGTLDKE							
160 170 180							

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1579

A DNA sequence (GBSx1673) was identified in *S.agalactiae* <SEQ ID 4877> which encodes the amino acid sequence <SEQ ID 4878>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -6.85 Transmembrane 86 - 102 (80 - 106)

----- Final Results -----

50 bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

55

-1764-

Example 1580

A DNA sequence (GBSx1674) was identified in *S.agalactiae* <SEQ ID 4879> which encodes the amino acid sequence <SEQ ID 4880>. Analysis of this protein sequence reveals the following:

Possible site: 47

5 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -3.61 Transmembrane 107 - 123 (96 - 124)
 INTEGRAL Likelihood = -1.86 Transmembrane 124 - 140 (124 - 142)
 INTEGRAL Likelihood = -1.38 Transmembrane 83 - 99 (83 - 100)
 10 INTEGRAL Likelihood = -1.12 Transmembrane 142 - 158 (142 - 160)

----- Final Results -----

15 bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9415> which encodes amino acid sequence <SEQ ID 9416> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAC76124 GB:AE000391 putative transport protein [Escherichia
 coli K12]
 Identities = 139/178 (78%), Positives = 159/178 (89%)
 Query: 1 MVGTMFLFVALVVPPIAFVMMRKNPYPLVLRCLKDSGITAFFTRSSAANIPVNMRLCEDL 60
 +VG ML VALVVPNP++ + +R+NP+PLVL CL++SG+ AFFTRSSAANIPVNM LCE L
 25 Sbjct: 222 LVGCMLLVALVVPNLLVWVKIRNPFPPLVLLCLRESGVYAFFTRSSAANIPVNMALCEKL 281
 Query: 61 GLDKDTYSVSIPLGAAINMAGAAITINILTLAAVNTLGITVDFTAFLLSVVAASACGA 120
 LD+DTYSVSIPLGA INMAGAAITI +LTLAAVNTLGI VD PTA LLSVVA++ ACGA
 30 Sbjct: 282 NLDKDTYSVSIPLGATINMAGAAITITVTLAAVNTLGIPVDLPTALLSVVASLCACGA 341
 Query: 121 SGVTGGSLLLIPVACSLFGISNDVAMQVVGFGVIGVQDSCETALNSSTDVLFATA 178
 SGV GGSLLLLIP+AC++FGISND+AMQVV VGFI+GV+QDSCETALNSSTDVLFATA A
 Sbjct: 342 SGVAGGSLLLLIPLACNMFGISNDIAMQVAVGFIIGVLQDSCETALNSSTDVLFATAA 399

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4881> which encodes the amino acid sequence <SEQ ID 4882>. Analysis of this protein sequence reveals the following:

Possible site: 58

40 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -13.69 Transmembrane 212 - 228 (202 - 239)
 INTEGRAL Likelihood = -7.38 Transmembrane 78 - 94 (74 - 108)
 INTEGRAL Likelihood = -6.53 Transmembrane 179 - 195 (175 - 200)
 INTEGRAL Likelihood = -6.10 Transmembrane 315 - 331 (312 - 341)
 INTEGRAL Likelihood = -5.36 Transmembrane 44 - 60 (42 - 61)
 45 INTEGRAL Likelihood = -4.41 Transmembrane 13 - 29 (11 - 41)
 INTEGRAL Likelihood = -3.19 Transmembrane 340 - 356 (333 - 358)
 INTEGRAL Likelihood = -3.08 Transmembrane 145 - 161 (144 - 162)
 INTEGRAL Likelihood = -0.90 Transmembrane 358 - 374 (358 - 376)

----- Final Results -----

50 bacterial membrane --- Certainty=0.6477(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 >GP:AAF95950 GB:AE004347 sodium/dicarboxylate symporter [Vibrio cholerae]
 Identities = 243/385 (63%), Positives = 299/385 (77%), Gaps = 2/385 (0%)
 Query: 9 VRVSLIKKIGIGVVGVMGLGILAPDLTG-FSILGKLFVGGKKAIAPLLVFALVSQAISHQ 67
 VR +L+ +I G+++G + +P+ ++G LFVG LKA+AP+LVF LV+ +I++Q

-1765-

Sbjct: 11 VRGNLVLQILAGILLGAAMATFSPEYAQKVGLIGNLFGALKAVAPVLVLFILVASSIANQ 70

Query: 68 KKGKQTNMTLIIIVLYLFGTFASALVAVLTAYLFPLTLVLNTPVNTELSPPQGVAEVFQSL 127
 KK + T M I+VLYLFGTF++AL AV+ ++LFP TLVL T +PPQG+AEV +L

5 Sbjct: 71 KKNQHTYMRPIVLYLFGTFSAALTAVILSFLFPTTLVLATGAEGA-TPPQGIAEVLNLT 129

Query: 128 LLKLVDNPINALATANYIGVLSWAIIFGLALKASKETKHLIKTAAEVT SQIVVWIINLA 187
 L KLVDNP++AL ANYIG+L+W + GLAL +S TK + + + SQIV +II LA

10 Sbjct: 130 LFKLVDNPVSALMNANYIGILAWGVGLGLALHHSSTTKAVFEDLSHGISQIVRFIIRLA 189

Query: 188 PIGIMSLVFTTISENGVILSDYAFILVLVGTMLFVALVNPPLIAVLITRQNPYPLVLR 247
 P GI LV +T + G L+ YA L+ VL+G M F+ALVVNP+I R+NP+PLVL+

Sbjct: 190 PFGIFGLVASTFATTGFDALAGYAQLLAVLLGAMAFIALVNPIMIVYKIRRNPFPLVLQ 249

15 Query: 248 CLRESGLTAFSTRSSAANIPVNMQLCQKIGLSKDTYSVSIPLGATINMGGAITINVLTL 307
 CLRESG+TAFSTRSSAANIPVNM LC+K+ L +DTYSVSIPLGATINM GAAITI VLTL

Sbjct: 250 CLRESGVTAFSTRSSAANIPVNMALCEKCLKLDEDTYSVSIPLGATINMAGAITITVLTL 309

Query: 308 AAVHTFGIPIDFLTALLSVVAAVSACGASGVAGGSLLLIPVACSLFGISNDLAMQVVG 367
 AAVHT GI +D +TALLSVVAAVSACGASGVAGGSLLLIP+AC LFGISND+AMQVV V

20 Sbjct: 310 AAVHTMGIEVDLMTALLSVVAAVSACGASGVAGGSLLLIPLACGLFGISNDIAMQVVAV 369

Query: 368 GFIVGVIQDSCETALNSSTDVLF 392
 GFI+GVIQDS ETALNSSTDVLF

25 Sbjct: 370 GFIIIGVIQDSAETALNSSTDVLF 394

An alignment of the GAS and GBS proteins is shown below.

Identities = 153/186 (82%), Positives = 172/186 (92%)

30 Query: 1 MVGTMLFVALVNPPIAFVMMRKNPYPLVLRCLKDSGITAFSTRSSAANIPVNMRLCEDL 60
 +VGTMLFVALVNP+IA ++ R+NPYPLVLRCL++SG+TAFSTRSSAANIPVNM+LC+ +

Sbjct: 217 LVGTMLFVALVNPPLIAVLITRQNPYPLVLRCLRESGLTAFSTRSSAANIPVNMQLCQKI 276

Query: 61 GLDKDTYSVSIPLGAANMAGAAITINILTLAAVNTLGITVDFTAFLLSVVAAVSACGA 120
 GL KDTYSVSIPLGA INM GAAITIN+LTLAAV+T GI +DF TA LLSVVAAVSACGA

35 Sbjct: 277 GLSKDTYSVSIPLGATINMGGAITINVLTLAAVHTFGIPIDFLTALLSVVAAVSACGA 336

Query: 121 SGVTGGSLLLIPVACSLFGISNDVAMQVVGFGFIVGVIQDSCETALNSSTDVLF 180
 SGV GGSLLIPVACSLFGISND+AMQVVGFGFIVGVIQDSCETALNSSTDVLF+AE

40 Sbjct: 337 SGVAGGSLLLIPVACSLFGISNDLAMQVVGFGFIVGVIQDSCETALNSSTDVLF 396

Query: 181 SVFGKK 186
 + + +K

45 Sbjct: 397 AFWKRRK 402

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1581

50 A DNA sequence (GBSx1675) was identified in *S.agalactiae* <SEQ ID 4883> which encodes the amino acid sequence <SEQ ID 4884>. This protein is predicted to be acid phosphatase. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2436(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1766-

A related GBS nucleic acid sequence <SEQ ID 9427> which encodes amino acid sequence <SEQ ID 9428> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:CAA73175 GB:Y12602 acid phosphatase [Streptococcus equisimilis]
    Identities = 167/251 (66%), Positives = 209/251 (82%)

    Query: 7  EQKTKFKNISLSSNKLAKENTMSVLWYQNSAEAKALYLQGYNVAKMKLDDWLQKPSEKP 66
          ++ K ++ S +L + ENTMSVLWYQ +AEAKALYLQGY +A +L + L + ++KP
    Sbjet: 34  KETVKQTKVTYSDEQLRSNENTMSVLWYQRAAEAKALYLQGYQLATDRLKNLQGLQATDKP 93

10  Query: 67  YSIILDLDETFLDNSPYQAKNIKDGSSFTPEWDKVVQKSAKAVAGAKEFLKYANEKGI 126
          YSI+LD+DETFLDNSPYQAKNI +G+SFTPEWD WVQKK AK VAGAKEFL++A++ G+
    Sbjet: 94  YSIVLDIDETFLDNSPYQAKNILEGTSFTPEWDVWVQKKEAKPVAGAKEFLQFADQNGV 153

15  Query: 127  KIYYVSDRTDAQVDATKENLEKEGIPVQGDHLLFLKKGMKSKESRRQAVQKDTNLIMLF 186
          +IYY+SDR +QVDAT ENL+KEGIPVQG+DHLLFL++G+KSKE+RRQ V++ TNLIMLF
    Sbjet: 154  QIYYISDRAVSQVDATMENLQKEGIPVQGRDHLLFLEEGVKSKEARRQKVKETTNLIMLF 213

20  Query: 187  GDNLVDFADFSKSSSTDREQLLTKLQSEFGSKFIVFPNPMYGSWESAIYQKHLDVQKQL 246
          GDNLVDFADFSK S DR LL++LQ EFG +FI+FPNPMYGSWESA+Y+G LD QL
    Sbjet: 214  GDNLVDFADFSKSEEDRTALLSELQEEFGRQFIIFPNPMYGSWESAVYKGDKLDASHQL 273

    Query: 247  KERQKMLHSYD 257
          KER+K L S++
25  Sbjet: 274  KERRKALESFE 284
  
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4885> which encodes the amino acid sequence <SEQ ID 4886>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 25

    >>> May be a lipoprotein

    ----- Final Results -----
35  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
  
```

The protein has homology with the following sequences in the databases:

```

40  >GP:CAA73175 GB:Y12602 acid phosphatase [Streptococcus equisimilis]
    Identities = 234/284 (82%), Positives = 261/284 (91%)

    Query: 1  MKSKKVVSVISLTLSLFLVTGCAKVDNNKSVNLKPKATKQTYNSYSDQLRSRENTMSVLW 60
          MK+K+V SVISL LSLFLVTGCA++D+ +VN K KQT +YSD+QLRS ENTMSVLW
    Sbjet: 1  MKTKQVASVISLALSLFLVTGCAQLDHKANVNSKETVKQTKVTYSDEQLRSNENTMSVLW 60

45  Query: 61  YQRAAETQALYLQGYQLATDRLKEQLNKPTDKPYSIVLDIDETFLDNSPYQAKNVLEGTG 120
          YQRAAE +ALYLQGYQLATDRLK QL + TDKPYSIVLDIDETFLDNSPYQAKN+LEGT
    Sbjet: 61  YQRAAEAKALYLQGYQLATDRLKNLQGLQATDKPYSIVLDIDETFLDNSPYQAKNILEGTS 120

50  Query: 121  FTPESWDYVWQKKEAKPVAGAKDFLQFADQNGVQIYYISDRSTTQVDATMENLQKEGIPV 180
          FTPESWD WVQKKEAKPVAGAK+FLQFADQNGVQIYYISDR+ +QVDATMENLQKEGIPV
    Sbjet: 121  FTPESWDVWQKKEAKPVAGAKEFLQFADQNGVQIYYISDRAVSQVDATMENLQKEGIPV 180

55  Query: 181  QGRDHLLFLEKGVKSKESSRRQKVKETTNTVMTLFGDNLDFADFSKKSQEDRTALLSDLQE 240
          QGRDHLLFLE+GVKSKE+RRQKVKETTNT+ MLFGDNL+DFADFSKKS+EDRTALLS+LQE
    Sbjet: 181  QGRDHLLFLEEGVKSKEARRQKVKETTNTLIMLFGDNLVDFADFSKKSSEEDRTALLSELQE 240

    Query: 241  EFGRRFIIFPNPMYGSWEGAIYKGEKLDVLKQLEERRKSLKSKF 284
          EFGR+FIIFPNPMYGSWE A+YKG+KLD QL+ERRK+L+SF+
60  Sbjet: 241  EFGROFIIFPNPMYGSWESAVYKGDKLDASHQLKERRKALESFE 284
  
```

An alignment of the GAS and GBS proteins is shown below.

-1767-

Identities = 166/247 (67%), Positives = 207/247 (83%)

Query: 10 TKFKNISLSSNKLAKENTMSVLWYQNSAEAKALYLQGYNVAKMKLDDWLQKPSEKPYSI 69
 TK S S ++L ++ENTMSVLWYQ +AE +ALYLQGY +A +L + L KP++KPYSI
 5 Sbjct: 37 TKQTYNSYSDDQLRSRENTMSVLWYQRAAETQALYLQGYQLATDRLKEQLNKPTDKPYSI 96

Query: 70 ILDLDETVDLNSPYQAKNIKDGSSFTPESWDKWVQKKSAAVAGAKEFLKYANEKGIKIY 129
 +LD+DETVDLNSPYQAKN+ +G+ FTPESWD WVQKK AK VAGAK+FL++A++ G++IY
 10 Sbjct: 97 VLDIDETVDLNSPYQAKNVLEGTGFTPESWDYWVQKKEAKPVAGAKDFLQFADQNGVQIY 156

Query: 130 YVSDRTDAQVDATKENLEKEGIPVQGDHLLFLKKGMSKESRRQAVQKDTNLIIMLFQDN 189
 Y+SDR+ QVDAT ENL+KEGIPVQG+DHLLFL+KG+KSKESRRQ V++ TN+ MLFGDN
 15 Sbjct: 157 YISDRSTTQVDATMENLQKEGIPVQGRDHLLFLEKGVKSKESRRQVKETTNTVMTLFGDN 216

Query: 190 LVDFADFSKSSSTDREQLLTKLQSEFGSKFIVFENPMYGSWESAIYQGKHLDVQKQLKER 249
 L+DFADFSK S DR LL+ LQ EFG +FI+FFNPMYGSWE AIY+G+ LDV KQL+ER
 20 Sbjct: 217 LLDFADFSKKSQEDRTALLSDLQEFGRRFIIFFNPMYGSWEGAIYKGEKLDVLKQLEER 276

Query: 250 QKMLHSY 256
 +K L S+
 20 Sbjct: 277 RKSLKSF 283

SEQ ID 9428 (GBS661) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 136 (lane 2 & 4; MW 61kDa + lane 3; MW 27kDa) and in Figure 186 (lane 11; MW 61kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 136 (lane 5-7; MW 25kDa).

GBS661-GST was purified as shown in Figure 237, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 1582

A DNA sequence (GBSx1676) was identified in *S.agalactiae* <SEQ ID 4887> which encodes the amino acid sequence <SEQ ID 4888>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

Possible site: 58
 35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3462(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4889> which encodes the amino acid sequence <SEQ ID 4890>. Analysis of this protein sequence reveals the following:

Possible site: 58
 45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3462(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 395/398 (99%), Positives = 398/398 (99%)

55 Query: 1 MAKLTIVKDVLDLKGKKVLRVDFNVPLKDGVIITNDNRITAAALPTIKYIIEQGGRAILFSHL 60

-1768-

MAKLTVKDVLDLKGKKVLRVDFNVPLKDGVTNDNRITAALPTIKYIEQGGRILFSLH
 Sbjct: 1 MAKLTVKDVLDLKGKKVLRVDFNVPLKDGVTNDNRITAALPTIKYIEQGGRILFSLH 60

Query: 61 GRVKEEADKEGKSLAPVAADLAAKLGQDVVFPGVTRGAKLEEAINALEDGQVLLVENTRF 120
 GRVKEEADKEGKSLAPVAADLAAKLGQDVVFPGVTRG+KLEEAINALEDGQVLLVENTRF
 Sbjct: 61 GRVKEEADKEGKSLAPVAADLAAKLGQDVVFPGVTRGSKLEEAINALEDGQVLLVENTRF 120

Query: 121 EDVDGKKESKNDDEELGKYWASLGDGIFVNDAFGTAHRAHASNVGISANVEKAVAGFLLEN 180
 EDVDGKKESKNDDEELGKYWASLGDGIFVNDAFGTAHRAHASNVGISANVEKAVAGFLLEN
 Sbjct: 121 EDVDGKKESKNDDEELGKYWASLGDGIFVNDAFGTAHRAHASNVGISANVEKAVAGFLLEN 180

Query: 181 EIAYIQEAVETPERPFVAILGGSKVS DKIGVIENLLEKADKVLIGGMYTFYKAQGIEI 240
 EIAYIQEAVETPERPFVAILGGSKVS DKIGVIENLLEKADKVLIGGMYTFYKAQGIEI
 Sbjct: 181 EIAYIQEAVETPERPFVAILGGSKVS DKIGVIENLLEKADKVLIGGMYTFYKAQGIEI 240

Query: 241 GNSLVEEDKLDVAKDLEKSNGKLILPVD SKANAFAGYTEVRDTEGEAVSEGFLGLDIG 300
 GNSLVEEDKLDVAKDLEKSNGKLILPVD SKANAFAGYTEVRDTEGEAVSEGFLGLDIG
 Sbjct: 241 GNSLVEEDKLDVAKDLEKSNGKLILPVD SKANAFAGYTEVRDTEGEAVSEGFLGLDIG 300

Query: 301 PKSI AKFDEALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIGGDSAAA 360
 PKSI A+FD+ALTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIGGDSAAA
 Sbjct: 301 PKSI AEFDAQLTGAKTVVWNGPMGVFENPDFQAGTIGVMDAIVKQPGVKSIIGGDSAAA 360

Query: 361 AINLGRADKFSWISTGGGASMELEGGKVLPGLAALTEK 398
 AINLGRADKFSWISTGGGASMELEGGKVLPGLAALTEK
 Sbjct: 361 AINLGRADKFSWISTGGGASMELEGGKVLPGLAALTEK 398

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 1583

A DNA sequence (GBSx1677) was identified in *S.galactiae* <SEQ ID 4891> which encodes the amino acid sequence <SEQ ID 4892>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.39	Transmembrane	97 - 113 (93 - 118)
INTEGRAL	Likelihood = -3.66	Transmembrane	25 - 41 (24 - 48)
INTEGRAL	Likelihood = -3.40	Transmembrane	121 - 137 (121 - 140)
INTEGRAL	Likelihood = -3.24	Transmembrane	72 - 88 (72 - 88)
INTEGRAL	Likelihood = -2.07	Transmembrane	143 - 159 (143 - 160)

----- Final Results -----

bacterial membrane	---	Certainty=0.4354(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4893> which encodes the amino acid sequence <SEQ ID 4894>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.23	Transmembrane	97 - 113 (93 - 118)
INTEGRAL	Likelihood = -7.17	Transmembrane	121 - 137 (119 - 140)
INTEGRAL	Likelihood = -4.19	Transmembrane	25 - 41 (24 - 48)
INTEGRAL	Likelihood = -3.24	Transmembrane	72 - 88 (72 - 88)
INTEGRAL	Likelihood = -2.55	Transmembrane	154 - 170 (154 - 170)

----- Final Results -----

bacterial membrane	---	Certainty=0.4291(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

-1769-

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 155/178 (87%), Positives = 169/178 (94%)

```

5      Query: 1  MKTLKKLLSNYKFDIKKFKLGMRTFKTGLSVFLVLLVFHFLFGWKGLQIGALTAVFSLRED 60
      Sbjct: 1  MKTLRKLSSNYKFDIKKFKLGMRTLKTGLSVFLVLLVFHFLFGWKGLQIGALTAVFSLRED 60

10     Query: 61  FDKSVHFGFSRIIGNSIGGLLSLVFFAFNEIFHQAFWVTLIVPICTMLCIMINVACNNK 120
      Sbjct: 61  FDKSVHFGFSRIIGNSIGGLLSLVFFAFNEIFHQAFWVTLIVPICTMLCIM+NVACNNK
      Sbjct: 61  FDKSVHFGFSRIIGNSIGGLLSLVFFAFNEIFHQAFWVTLIVPICTMLCIMVNVACNNK 120

15     Query: 121 SGIIGGTAALLIITLSIPSGETILYVFARIFETFCGVFIAMMVNTDIEILRKKLNK 178
      Sbjct: 121 SGIIGAVAALLIITLSIPTGTFTIYVTSRVFETFCGVFVAILVNTDVELIKNKWFNKK 178

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 1584

A DNA sequence (GBSx1678) was identified in *S.agalactiae* <SEQ ID 4895> which encodes the amino acid sequence <SEQ ID 4896>. This protein is predicted to be regulatory protein glr (glrR). Analysis of this protein sequence reveals the following:

```

25     Possible site: 17
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
30     bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

      >GP:BAA00402 GB:D00513 ORF129 [Bacillus cereus]
      Identities = 59/123 (47%), Positives = 89/123 (71%), Gaps = 5/123 (4%)

35     Query: 4  RELRRITMAVFPIGAVMKLTDLTARQIRYYEDQGLITPERTEGNRRMFSLNMDRLLEIKD 63
      Sbjct: 2  KEDRRSAPLFPPIGIVMDLTQLSARQIRYYEEHNLVSPTRTKGNRRLFNFNDVDKLLEIKD 61

40     Query: 64  FISDGLHISDIKNEYMQRQH-----KSKEKQKSLSDAEVRRLQDELNRNQGRFSSPSQHI 118
      Sbjct: 62  LLDQGLNMAGIKQVLLMKENQTEAVKVKEETKEISKTELRLRDELQHTGRFNRTSLRQ 121

45     Query: 119 GNM 121
      Sbjct: 122 GDI 124

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4897> which encodes the amino acid sequence <SEQ ID 4898>. Analysis of this protein sequence reveals the following:

```

50     Possible site: 20
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
55     bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

-1770-

The protein has homology with the following sequences in the databases:

```

>GP:BAA00402 GB:D00513 ORF129 [Bacillus cereus]
Identities = 59/122 (48%), Positives = 83/122 (67%), Gaps = 5/122 (4%)

5   Query: 4   KELRRSMAVFPPIGTVMTLTDL SARQIRYYEDQGLIKPERTQGNRRMFSLNMDRLLEIKD 63
      KE RRS +FPIG VM LT LSARQIRYYE+ L+ P RT+GNRR+FS ND+D+LLEIKD
      Sbjct: 2   KEDRRSAPLFFPIGIVMDLTQLSARQIRYYEEHNLVSPTRTKGNRRLFSFNDVDKLEIKD 61

10  Query: 64   FLSEGLNIAAIKREYVERQG-----KLMQKQKALTDADVRRILHDEMLTQSGFSTPSQHI 118
      L +GLN+A IK+ + ++ K+ ++ K ++ ++R+IL DE+ F+ S
      Sbjct: 62   LLDQGLNMAGIKQVLLMKENQTEAVKVKEETKETSKTELKILRDELQHTGRFNRTSLRQ 121

      Query: 119 GN 120
      G+
15  Sbjct: 122 GD 123

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 90/123 (73%), Positives = 108/123 (87%)

20  Query: 1   MKERELRRTMAVFPPIGAVMKLTDLTARQIRYYEDQGLITPERTEGNRRMFSLNMDRLLE 60
      MKE+ELRR+MAVFPPIG VM LTDL+ARQIRYYEDQGLI PERT+GNRRMFSLNMDRLLE
      Sbjct: 1   MKEKELRRSMAVFPPIGTVMTLTDL SARQIRYYEDQGLIKPERTQGNRRMFSLNMDRLLE 60

25  Query: 61   IKDFISDGLHISDIKNEYMQRQHSKEKQKSLSDAEVRRLLQDELNRNQGRFSSPSQHIGN 120
      IKDF+S+GL+I+ IK EY++RQ K +KQK+L+DA+VRR+L DE+ Q FS+PSQHIGN
      Sbjct: 61   IKDFLSEGLNIAAIKREYVERQGKLMQKQKALTDADVRRILHDEMLTQSGFSTPSQHIGN 120

      Query: 121 MHL 123
      +
30  Sbjct: 121 FRI 123

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1585

35 A DNA sequence (GBSx1679) was identified in *S.agalactiae* <SEQ ID 4899> which encodes the amino acid sequence <SEQ ID 4900>. This protein is predicted to be glutamine synthetase (glnA). Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

40  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2157(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4901> which encodes the amino acid sequence <SEQ ID 4902>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence
50  INTEGRAL Likelihood = -0.00 Transmembrane 347 - 363 ( 347 - 363)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
55  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-1771-

Identities = 392/448 (87%), Positives = 421/448 (93%)

Query: 1 MTITAEDIRREVKEKNVTFRLMFTDILGVMKNVEIPATDEQLDKVLSNKAMFDGSSIEG 60
 M IT DIRREVKEKNVTFRLMFTDI+GVMKNVEIPAT EQLDKVLSNK MFDGSSIEG
 5 Sbjet: 1 MAITVADIRREVKEKNVTFRLMFTDIMGVMKNVEIPATKEQLDKVLSNKVMFDGSSIEG 60

Query: 61 FVRINESDMYLYPDLDTWIVFPWGDENGAVAGLICDIYTAEGEPFAGDPRGNLKRNMKRM 120
 FVRINESDMYLYPDLDTWIVFPWGDENGAVAGLICDIYTAEG+PFAGDPRGNLKR +K M
 10 Sbjet: 61 FVRINESDMYLYPDLDTWIVFPWGDENGAVAGLICDIYTAEGKPFAGDPRGNLKRALKHM 120

Query: 121 QEMGYKSFNLGPPEFFFLFKMDENGNPFLDNDKGGYFDLAPIDLADNTRREIVNVLITQM 180
 E+GYKSFNLGPPEFFFLFKMD+ GNPTL+VND GGYFDLAP DLADNTRREIVN+LT+M
 Sbjet: 121 NEIGYKSFNLGPPEFFFLFKMDKGNPTLEVNDNGGYFDLAPIDLADNTRREIVNILTQM 180

Query: 181 GFEVEASHHEVAVGQHEIDFKYDDVLKACDNIQLFKLVVKTARKHGLYATFMAKPKFGI 240
 GFEVEASHHEVAVGQHEIDFKY DVLKACDNIQ+FKLVVKTAR+HGLYATFMAKPKFGI
 15 Sbjet: 181 GFEVEASHHEVAVGQHEIDFKYADVLKACDNIQIFKLVVKTAREHGLYATFMAKPKFGI 240

Query: 241 NGSGMHCNMSLFDNNEGNAFFDPEDPRGMQLSEDAYYFLGGLMKHAYNYTAIINPTVNSY 300
 GSGMHCNMSLFDN+GNNAF+D D RGMQLSEDAYYFLGGLMKHAYNYTAI NPTVNSY
 20 Sbjet: 241 AGSGMHCNMSLFDNQCNGNAFYDEADKRGMLSEDAYYFLGGLMKHAYNYTAITNPTVNSY 300

Query: 301 KRLVPGYEAPVYVAVAGNRNSPLIRVPASRGMGTRELRSDPTANPYLALSVLLGSGLE 360
 KRLVPGYEAPVYVAVAG NRSPILIRVPASRGMGTRELRSDPTANPYLAL+VLL +GL+
 25 Sbjet: 301 KRLVPGYEAPVYVAVAGSNRNSPLIRVPASRGMGTRELRSDPTANPYLALAVLLEAGLD 360

Query: 361 GIENKIEAPEPIETNIYAMTVEERRQAGIVDLPSTLHNALEALEBDEVVKAALGTHIYTN 420
 GI NKIEAPEP+E NIY MT+EER +AGI+DLPSTLHNAL+AL++D+VV+ ALG HIYTN
 30 Sbjet: 361 GIINKIEAPEPVEANIYTMTEERNEAGIIDLPSTLHNALKALQKDDVVQKALGYHIYTN 420

Query: 421 FLDKRIEWASYATYVSQWEIDNYLDLY 448
 FL+AKRIEW+SYAT+VSQWEID+Y+ Y
 Sbjet: 421 FLEAKRIEWSSYATFVSQWEIDHYIHN 448

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1586

A DNA sequence (GBSx1680) was identified in *S.agalactiae* <SEQ ID 4903> which encodes the amino acid sequence <SEQ ID 4904>. This protein is predicted to be SceB precursor. Analysis of this protein
 40 sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:CAA66624 GB:X97985 ORF1 [Staphylococcus aureus]
 Identities = 44/119 (36%), Positives = 66/119 (54%), Gaps = 4/119 (3%)

Query: 26 SFASTNADANTYNYAVDVDYLASAEETIAQHPA-SNTFPLGQCTWGVKE-MATWAGNWWG 83
 S AS + +N + ++ I+ + + SN + GQCT+ V + + G+ WG
 55 Sbjet: 117 SGASYSTTSNNVHVTTTAAPSSNGRSISNGYASGSNLYTSGQCTYYVFDVRVGGKIGSTWG 176

Query: 84 NGGDWAASAASADYTVGTQPRVGSIVCWTDGSYGHVAVYTAVDPVTNKIQVLESNYAGH 142
 N +WA +AAS+ YTV P+VG+I+ T G YGHVAYV V+ ++V E NY GH
 60 Sbjet: 177 NASNWNAAAASSGYTVNNTPKVGAIMQTQGGYGHVAVVEGVNS-NGSVRVSEMNY-GH 233

-1772-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1013> which encodes the amino acid sequence <SEQ ID 1014>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 60/115 (52%), Positives = 81/115 (70%), Gaps = 7/115 (6%)

Query: 55 AHPASNTFPLGQCTWGVKEMATWAGNWWGNGGDWAASAASADYTVGTQPRVGSIVCWTGD 114
++ +SNT+P+GQCTWG K +A WAGN WNGG WA SA +A Y G+ P VG+I W DG
Sbjct: 291 SYDSSNTYPVGQCTWGAKSLAPWAGNWWGNGGQWYSAQAAGYRTGSTPMVGAIAVWWDG 350

Query: 115 SYGHVAVYTAVDPTNKKIQVLESNYAGHQWIDNYRGWFDPTQNTVTPGVVSYIYPN 169
YGHVA V V ++ I+V+ESNY+G Q+I ++RGWF+P V++IYP+
Sbjct: 351 GYGHVAVVVEVQSASS-IRVMESNYSGRQYIADHRGWFNPTG-----VTFTIYPH 398

A related GBS gene <SEQ ID 8859> and protein <SEQ ID 8860> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8

McG: Discrim Score: 5.85

GvH: Signal Score (-7.5): 3.11

Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 0 value: 6.74 threshold: 0.0

PERIPHERAL Likelihood = 6.74 115

modified ALOM score: -1.85

*** Reasoning Step: 3

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

37.5/56.7% over 200aa

Staphylococcus aureus

GP|1340128| ORF1 Insert characterized

ORF00255(376 - 726 of 1107)

GP|1340128|emb|CAA66624.1||X97985(33 - 233 of 255) ORF1 {Staphylococcus aureus}

%Match = 9.0

%Identity = 37.5 %Similarity = 56.7

Matches = 45 Mismatches = 47 Conservative Sub.s = 23

294 324 354 384 414
SVIWI**TRSHQMEENMNIKQLKSKTMLGTVLVSFAFSFASTNADANTYNYAVDVD-----
: : | : : | : : | : :
MKKIVTATITATAGLATIAFAGHDAQAEEQNNNGYNSNDAQSYSTYTTIDAQGNHYHTWIGNWNPSQLTQNN-----
10 20 30 40 50 60 70

462 489 516 546 576 606
-----YLASAEETIAQAHPA-SNTFPLGQCTWGV-KEMATWAGNWWGNGGDWAASAASADYTVGTQ
: : | : : | : | : | : : | : : | : : | : :
GSGASYSTTSNNVHVTTTAAPSNGRSISNGYASGSNLYTSGQCTYYVFDVRVGGKIGSTWGNASNWANAASSGYTVNNT
130 140 150 160 170 180 190

636 666 696 726 756 786 816 846
PRVGSIVCWTGDSYGHVAVYTAVDPTVTKIQVLESNYAGHQWIDNRYGFDPQNTVTTPGVVSIIYPN*SIKNSSHRRYKS
|::||: | ||||| | : :: | | | | : :
5 PKVGAIMQTQGYIYGHVAYVEGVNS-NGSVRVSEMN-Y-GHGAGVVTSRTISANQAGSYNFIIH
 210 220 230 240 250

GBS30-GST was purified as shown in Figure 193, lane 8.

Example 1587

```
Possible site: 14
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -3.93    Transmembrane    2 - 18 ( 1 - 18)
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 **Example 1588**

A DNA sequence (GBSx1682) was identified in *S.agalactiae* <SEQ ID 4907> which encodes the amino acid sequence <SEQ ID 4908>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have no N-terminal signal sequence

35
----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2160(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06381 GB:AP001516 unknown conserved protein [Bacillus halodurans]
Identities = 353/550 (64%), Positives = 443/550 (80%)

45	Query: 6	LKPEEVGVYAIGGLGEIGKNTYGIEYQDEIITVDAGIKFPEDDLLGIDYVIPDYSYIVEN 65
		LK + VYA+GGLGEIGKNTY +++QDEII++DAGIKFPED+LGIDYVIPDYSY+V+N
	Sbjct: 4	LKNNQTAVYALGGLGEIGKNTYAVQFQDEIILIDAGIKFPEDDELLGIDYVIPDYSYLVKN 63
50	Query: 66	IDRIKALVITHGHEDHIGGIPFLKQANLPYAGPLALALIKGKLEEHGLLRDATLYEIH 125
		++IK L ITHGHEDHIGGIP+LL++ N+PIY G LAL L++GKLEEHGLLR A L++I
	Sbjct: 64	ENKIKGLFVITHGHEDHIGGIPYLLREVNIPYGGKLAGLLRGKIPKLEEHGLLRKAKLHDIO 123

-1774-

5 Query: 126 ANELTFKNLSVTFRRTHSIP+ GIV+ TP G ++ TGDFKFDFTPVGEPA+L +MA
Sbjct: 124 EDDIIKFAKTSVSFFRTHSIPDSYGVVKTTPGNIVHTGDFKFDFTPVGEPA+LTKMAK 183

10 Query: 186 LGEDGVLCLLSDSTNAEVPFTFTNSEKIVGQSIMKIEGIEGRIIFASFASNI+RLQQAEE 245
+GE+GVLCLLSDSTN+E+P FT SE+ VG+SI I +EGRIIFA+FASNI RLQQA E
Sbjct: 184 IGEEGVLCLLSDSTNSEIPEFTMSERKVGESIDHIFRRVEGRIIFATFASNIHRLQQAEE 243

15 Query: 246 AAVKTGRKIAVFGSRMEKAI+VNGIELGYIKVPKGT+IEPSELKNLHASEVLMCTGSQGE 305
+AV+ GRK+AVFGSRME AI G ELGYIK PK TIEP++L L +EV+I+CTGSQGE
Sbjct: 244 SAVRYGRKVA+VFGSRMESAINIGQELGYIKAPKNT+IEPNQLNKLDPNEVMILCTGSQGE 303

20 Query: 306 SMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHGKINNI 365
MAAL+R+A GTHRQ+ + PGDTVIFSSSPIPGNT SV+K IN + +AG +VIHG +N+I
Sbjct: 304 PMAALSRVAFGTHRQIQIIPGDTVIFSSSPIPGNTLSVSKTINQLYKAGANVIHGSLNDI 363

25 Query: 366 HTSGHGGQEQKLMRLIKPKYFMPVHGEYRMQKVHAGLAVDTGIPKENIFIMENGDVLA 425
HTSGHGGQ+EQKLMRLIKPKYFMP+HGEYRM K+H LA D G+P EN FIM+NGDVLA
Sbjct: 364 HTSGHGGQEQKLMRLIKPKYFMPHGEYRMKMH+KLAEDCGVPAENCFIMDNGDVLA 423

30 Query: 426 LTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRHDLSEDGVVLAVATVDFDSKMILAG 485
L D A IAG + +YVDGNGIGDIG VLRDR LSE+G+V+ V +++ + AG
Sbjct: 424 LHPDEAGIAGKIPSGSVYVDGNGIGDIGNIVLRDRRLSEGLVVVVVSLNMKEYKV+TAG 483

35 Query: 486 PDILSRGFIYMRSGDLIRESOHILFNIRIALKNK+DASIQSVNGAIVNALRPFLYEKTE 545
PD++SRGF+YMRSGDLI+E+Q +L N ++ ++ K + I + L PFLY++T+
Sbjct: 484 PDLISRGFVYMRSGDLIQAORLLANHLQEVMERKTNQWSEIKNEITDVLGPFLYDRTK 543

40 Query: 546 REPILIPMVL 555
R+P+I+P+++
Sbjct: 544 RKPMILPIIM 553

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4909> which encodes the amino acid
35 sequence <SEQ ID 4910>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.11 Transmembrane 468 - 484 (468 - 484)

----- Final Results -----

bacterial membrane --- Certainty=0.1044 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB06381 GB:AP001516 unknown conserved protein [Bacillus halodurans]
Identities = 353/550 (64%), Positives = 444/550 (80%)

50 Query: 6 LKPNEVG+VFAIGGLGEIGKNTY+IEYQDEII+VDAGIKFPEDDLLGIDYVIPDYSYIVDN 65
LK N+ V+A+GGLGEIGKNTY +++QDEII++DAGIKFPED+LLGIDYVIPDYSY+V N
Sbjct: 4 LKNNQFAVYALGGLGEIGKNTYAVQFQDEIILIDAGIKFPEDELLGIDYVIPDYSYLVKN 63

55 Query: 66 LDRVKALVITHGHEDHIGGIPFLKQANIPIYAGPLALALIRKLEEHGLWREATVYEIN 125
+++K L ITHGHEDHIGGIP+LL++ NIPIY G LAL L+RGKLEEHGL R+A +++I
Sbjct: 64 ENKIKGLFITHGHEDHIGGIPYLLREVNIPYGGKLALGLLRKLEEHGLLRKAKLHDIQ 123

60 Query: 126 HNELTFKNLSVTFRRTHSIP+ GIV+ TP G I+ TGDFKFDFTPVGEPA+L +MA
+ + F SV+FF+TTHSIP+ GIV+ TP G I+ TGDFKFDFTPVGEPA+L +MA
Sbjct: 124 EDDIIKFAKTSVSFFRTHSIPDSYGVVKTTPGNIVHTGDFKFDFTPVGEPA+LTKMAK 183

65 Query: 186 LGEEGVLCLLSDSTNAEIPFTFTNSEKIVGQSILKIEGIEGRIIFASFASNI+RLQQAEE 245
+GEEGVLCLLSDSTN+EIP FT SE+ VG+SI I + GRIIFA+FASNI+RLQQA E
Sbjct: 184 IGEEGVLCLLSDSTNSEIPEFTMSERKVGESIDHIFRRVEGRIIFATFASNIHRLQQAEE 243

-1775-

- Query: 246 AAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPKGTFTIEPSELKNLHASEVLIMCTGSQGE 305
 +AV+ GRK+AVFGRSME AI G ELGYIK PK FTIEP++L L +EV+I+CTGSQGE
 Sbjct: 244 SAVRYGRKVAVFGRSMESAINIGQELGYIKAPKNTFTIEPNQLNKLDPNEVMILCTGSQGE 303
- 5 Query: 306 SMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHGKVNNI 365
 MAAL+R+A GTHRQ+ + PGDTVIFSSSPIPGNT SV+K IN + +AG +VIHG +N+I
 Sbjct: 304 PMAALSRVAFGTHRQIQIIPGDTVIFSSSPIPGNTLSVSKTINQLYKAGANVIHGSLNDI 363
- 10 Query: 366 HTSGHGGQEQKLMLSLIKPKYFMPVHGEYRMQKVHAGLAMDIGIPKENIFIMENGVDVLA 425
 HTSGHGGQ+EQKLML LIKPKYFMP+HGEYRM K+H LA D G+P EN FIM+NGVDVLA
 Sbjct: 364 HTSGHGGQEQKLMLRLIKPKYFMP+HGEYRMLKMHTKLAEDCGVPAENCFIMDNGVDVLA 423
- 15 Query: 426 LTSDSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRDLSEDGVVLAVATVDFNTQMILAG 485
 L D A IAG + +YVDGNGIGDIG VLRDRR LSE+G+V+ V +++ + AG
 Sbjct: 424 LHPDEAGIAGKIPSGSVYVDGNGIGDIGNTVLRDRRILSEEGLVVVVVSLNMKEYKVVTAG 483
- 20 Query: 486 PDILSRGFIYMRESGDLIRESQVLFNAIRIALKNKDASIQSVNGAIVNALRPFLEYKTE 545
 PD++SRGF+YMRESGDLI+E+QR+L N ++ ++ K + I + L PFLY++T+
 Sbjct: 484 PDLISRGFVYMRESGDLIQEAQRLLANHLQEVMERKTNQWSEIKNEITDVLGPFLEYDRTK 543
- Query: 546 REPIIIPMVL 555
 R+P+I+P+++
 Sbjct: 544 RKPMLPIIM 553
- 25 An alignment of the GAS and GBS proteins is shown below.
 Identities = 523/559 (93%), Positives = 550/559 (97%)
- Query: 1 MSNINLKPEEVGVYAIGGLGEIGKNTYGYIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYS 60
 M+NI+LKP EVGV+AIIGLGEIGKNTYGYIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYS
 30 Sbjct: 1 MTNISLKPNEVGVAIGGLGEIGKNTYGYIEYQDEIIIVDAGIKFPEDDLLGIDYVIPDYS 60
- Query: 61 YIVENIDRIKALVITHGHEDHIGGIPFLKQANLPIYAGPLALALIKGLEEHGLLRDAT 120
 YIV+N+DR+KALVITHGHEDHIGGIPFLKQAN+PIYAGPLALALI+KLEEHGL R+AT
 35 Sbjct: 61 YIVDNLDVRKALVITHGHEDHIGGIPFLKQANLPIYAGPLALALIRKLEEHGLWREAT 120
- Query: 121 LYEIHANTELTFKNLSVTFRTTHSIEPLGIVIHTPQGKVICTGDFKFDFTPVGEPADL 180
 +YEI+ NTELTFFKN+SVTFF+TTHSIEP+GIVIHTPQGK+ICTGDFKFDFTPVG+PADL
 Sbjct: 121 VYEINHNTLTFKNMSVTFKTTTHSIEPVGIVIHTPQGKIICTGDFKFDFTPVGDPADL 180
- 40 Query: 181 HRMAALGEDGVLCCLSDSTNAEVPTFTNSEKIVGQSIKIIIEGRIIFASFASNIYRL 240
 RMAALGE+GVLCCLSDSTNAE+PTFTNSEK+VGQSI+KIIIEGRIIFASFASNI+RL
 Sbjct: 181 QRMALGEEGVLCCLSDSTNAEIPFTFTNSEKVVQSIKIIIEGRIIFASFASNIYRL 240
- 45 Query: 241 QQAEEAAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPKGTFTIEPSELKNLHASEVLIMCT 300
 QQAEEAAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPKGTFTIEPSELKNLHASEVLIMCT
 Sbjct: 241 QQAEEAAVKTGRKIAVFGRSMEKAIVNGIELGYIKVPKGTFTIEPSELKNLHASEVLIMCT 300
- 50 Query: 301 GSQGESMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHG 360
 GSQGESMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHG
 Sbjct: 301 GSQGESMAALARIANGTHRQVTLQPGDTVIFSSSPIPGNTTSVNKLINTIQEAGVDVIHG 360
- 55 Query: 361 KNNIHTSGHGGQEQKLMLRLIKPKYFMPVHGEYRMQKVHAGLAVDTGIPKENIFIMEN 420
 K+NNIHTSGHGGQEQKLML LIKPKYFMPVHGEYRMQKVHAGLA+D GIPKENIFIMEN
 Sbjct: 361 KVNNIHTSGHGGQEQKLMLSLIKPKYFMPVHGEYRMQKVHAGLAMDIGIPKENIFIMEN 420
- 60 Query: 421 GDVLALTSARSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRDLSEDGVVLAVATVDFDSK 480
 GDVLALTSARSARIAGHFNAQDIYVDGNGIGDIGAAVLRDR DLSEDGVVLAVATVDF+++
 Sbjct: 421 GDVLALTSARSARIAGHFNAQDIYVDGNGIGDIGAAVLRDRDLSEDGVVLAVATVDFNTQ 480
- 65 Query: 481 MILAGPDILSRGFIYMRESGDLIRESQHILFNAIRIALKNKDASIQSVNGAIVNALRPF 540
 MILAGPDILSRGFIYMRESGDLIRESQ +LFNAIRIALKNKDASIQSVNGAIVNALRPF
 Sbjct: 481 MILAGPDILSRGFIYMRESGDLIRESQVLFNAIRIALKNKDASIQSVNGAIVNALRPF 540
- Query: 541 YEKTEREPIIIPMVLTPDK 559
 YEKTEREPIIIPMVLTPDK
 Sbjct: 541 YEKTEREPIIIPMVLTPDK 559

-1776-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1589

A DNA sequence (GBSx1683) was identified in *S.agalactiae* <SEQ ID 4911> which encodes the amino acid sequence <SEQ ID 4912>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.2932(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAB13327 GB:Z99111 ykzG [Bacillus subtilis]
 Identities = 27/75 (36%), Positives = 44/75 (58%), Gaps = 7/75 (9%)

 Query: 1 MIYKVFYQETKERNPRREQTKTLYVTIDAANELEGRIAARKLVEENTAYNIEFIELLSDK 60
 MIYKVFYQE + P RE+T +LY+ + ++ ++ +K +NIEFI +
20 Sbjct: 1 MIYKVFYQEKADVPVREKTDLSLYIEGVSEKDVRTKLKEK-----FNIEFITPVDGA 53

 Query: 61 HLEYEKETGVFELTE 75
 LEYE+++ F++ E
25 Sbjct: 54 FLEYEQQSENFKVLE 68

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4913> which encodes the amino acid sequence <SEQ ID 4914>. Analysis of this protein sequence reveals the following:

Possible site: 32
30 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3428(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 60/76 (78%), Positives = 70/76 (91%)

40 Query: 1 MIYKVFYQETKERNPRREQTKTLYVTIDAANELEGRIAARKLVEENTAYNIEFIELLSDK 60
 MIYKVFYQETK+++PRRE TK LY+ IDA +EL+GRI AR+LVE+NT YN+EFIELLSDK
 Sbjct: 1 MIYKVFYQETKDQSPRESTKALYLNIDATDELGDRIKARRLVEDNTYYNVEFIELLSDK 60

 Query: 61 HLEYEKETGVFELTEF 76
 HL+YEKETGVFELTEF
45 Sbjct: 61 HLDYEKETGVFELTEF 76

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1590

50 A DNA sequence (GBSx1684) was identified in *S.agalactiae* <SEQ ID 4915> which encodes the amino acid sequence <SEQ ID 4916>. This protein is predicted to be glycoprotein endopeptidase. Analysis of this protein sequence reveals the following:

Possible site: 13

-1777-

>>> Seems to have no N-terminal signal sequence (or aa 1-17)

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.0430(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA76861 GB:Y17797 hypothetical protein [Enterococcus faecalis]
 Identities = 94/182 (51%), Positives = 127/182 (69%), Gaps = 6/182 (3%)

15 Query: 2 MKVLAFDTSKALSVAVLNNMECLATVTINIKKNHSINLMPAIDFLMQSIDLEPQDLDR 61
 +++LA DTS++ LS+AV N + L + T +K+NHS+ LMPAID+LM ++L P +DR
 Sbjct: 13 VRILAIDTSNQTLSIAVCENQKILGSYTATVKKRNHSLTLPAYDYLMSQLNLAPTADR 72

 Query: 62 VVAEGPGSYTGRLRVAVATAKMLAYTLKIDLVGVSSLYAL-TNGFSENDLLVPLIDARRNN 120
 VVAEGPGSYTGRLR+ V TAK LAYTLK +LVG+SSL AL N + L+VPL DARR N
 Sbjct: 73 VVAEGPGSYTGRLRLGVTTAKTLAYTLKKELVGISSLQALAANCVGQTGLIVPLFDARRKN 132

20 Query: 121 VYVGIFYQNGDTV----KPDCHTSLEEVLQEVGNKANVHFVGE-VAAFFDQIKKALPHAKI 175
 VY G Y+ D V PD H SL E+L+++ N+ N+ FVGE V F ++I + +PH +I
 Sbjct: 133 VYAGAYRFVDGVWQNELPDQHSRLLELLEQLKNEPNLFFVGEDVEKFTEETAQIIPHGEI 192

25 Query: 176 TE 177
 +
 Sbjct: 193 CD 194

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4917> which encodes the amino acid sequence <SEQ ID 4918>. Analysis of this protein sequence reveals the following:

30 Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 99 - 115 (99 - 115)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9159> which encodes the amino acid sequence <SEQ ID 9160>. Analysis of this protein sequence reveals the following:

40 Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 88 - 104 (88 - 104)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

 Identities = 134/232 (57%), Positives = 172/232 (73%), Gaps = 3/232 (1%)

55 Query: 2 MKVLAFDTSKALSVAVLNNMECLATVTINIKKNHSINLMPAIDFLMQSIDLEPQDLDR 61
 MK LAFDTS+K LS+A+L++ LA +T+NI+K HS++LMPAIDFLM DL+PQDL+RI
 Sbjct: 12 MKTLAFDTSNKRTLSLAILDDETLLADMTLNIQKKHSVSLMPAIDFLMTCTDLKPQDLERI 71

 Query: 62 VVAEGPGSYTGRLRVAVATAKMLAYTLKIDLVGVSSLYALTNGFSE---NDLLVPLIDARR 118
 VVA+GPGSYTGRLRVAVATAK LAY+L I LVG+SSLYAL + N L+VPLIDARR
 Sbjct: 72 VVAKGPGSYTGRLRVAVATAKTLAYSLNIALVGISSLYALAASTCKQYPNTLVVPLIDARR 131

60 Query: 119 NNVVYVGIFYQNGDTVKPDCHTSLEEVLQEVGNKANVHFVGEVAAFFDQIKKALPHAKITET 178
 N YVG+Y+ G +V P H SLE +++++ + + FVGE A F ++I+K LP A + T

-1778-

Sbjct: 132 QNAYVGYRQGKSVMPQAHASLEVIIEQLVEEGQLIFVGETAPFAEKIQKKLPQAILLPT 191

Query: 179 LPCAVAIGRKQKMKSVNVDVFPYLRVEAEENWLKXHCETNTEEYIKRV 230

LP A G GQ + NVDAFVP+YLKRVEAEENWLK++ + Y+KR+

Sbjct: 192 LPSAYECGLLGQSLAPENVDAFVPQYLKRVEAEENWLKDNEIKDDSHYVKRI 243

SEQ ID 4916 (GBS69) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 9; MW 28.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 4; MW 53.9kDa).

- 10 The GBS69-GST fusion product was purified (Figure 197, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 285), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 1591

A DNA sequence (GBSx1685) was identified in *S.agalactiae* <SEQ ID 4919> which encodes the amino acid sequence <SEQ ID 4920>. This protein is predicted to be ribosomal-protein-alanine acetyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 22

- 20 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

- 25 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10137> which encodes amino acid sequence <SEQ ID 10138> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- 30 >GP:AAC06803 GB:AE000696 ribosomal-protein-alanine acetyltransferase
[Aquifex aeolicus]
Identities = 44/141 (31%), Positives = 74/141 (52%), Gaps = 8/141 (5%)

Query: 9 LREFEMESSEQALAIWSVLSVDYDKSPWSLSQISEDLKKDSTDYFFVYNDGEVIGFLALQ 68

35 +RE E E E+ ++ + + + WS +D + + F + DG+V+G++

Sbjct: 4 VREMEREDVER---VYEINRESFTTDAWSRFSFEKDFENKFSRRFVLEEDGKVVGYVIFW 60

Query: 69 QLVGEVBITNIAVKKNYQKGYAYQLM-----SMIADIEVPVFLEVRYSNIVAQKLYERCG 124

40 + E I A+ Y+GKGY +L+ S + D V L+VR SN+ A LY++ G

Sbjct: 61 VVKEEATIMTFAIAPGYRGKGYGEKLLREALSRLGDKVKRVVLDVRKSNLRNLYKKLIG 120

Query: 125 FVVLKRKNVYHDPEDAIVM 145

F V+ +RK YY D E+A++M

45 Sbjct: 121 FKVVTERKGYYSYG-ENALLM 140

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4921> which encodes the amino acid sequence <SEQ ID 4922>. Analysis of this protein sequence reveals the following:

Possible site: 35

- 50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3800(Affirmative) < succ>

-1779-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 65/140 (46%), Positives = 96/140 (68%), Gaps = 1/140 (0%)

Query: 9 LREFEMES-SEQALAIWSVLSDVYDKSPWSLSQISEDLKKDSTDYFFVYNDGEVIGFLAL 67
 L E M++ EQA I+ +L VY SPW+L Q+ D+++D TDYF +Y+ +++GFLA+
 10 Sbjct: 6 LSESNMKTVEEQAKNIYQLLEMVYGTSPWTLEQVLIDIRRDQTDYFLLYDHDKLLGLAI 65

Query: 68 QQLVGEVEITNIAVKKNYQKGAYQLMSMIADIEVPVFLEVRYSNIVAQKLYERCGFVV 127
 Q L GEVE+T IA+ ++Q G A QLM+ + IE +FLEVR SN AQ LY++ GF
 10 Sbjct: 66 QDLAGEVEMTQIAILPSHQELGLASQLMTHLDSIESDIFLEVRESNHRAQGLYQKFGFKF 125

15 Query: 128 LRKRKNYYHDPIDAIVMRK 147
 + KR +YY +PIE A++M++
 Sbjct: 126 IGRPDYRNPJETALLMKR 145

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1592

A DNA sequence (GBSx1686) was identified in *S.agalactiae* <SEQ ID 4923> which encodes the amino acid sequence <SEQ ID 4924>. Analysis of this protein sequence reveals the following:

25 Possible site: 21
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0334 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1593

A DNA sequence (GBSx1687) was identified in *S.agalactiae* <SEQ ID 4925> which encodes the amino acid sequence <SEQ ID 4926>. Analysis of this protein sequence reveals the following:

40 Possible site: 38
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.75 Transmembrane 86 - 102 (86 - 104)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.1702 (Affirmative) < succ>
 45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:BAB04267 GB:AP001508 glycoprotein endopeptidase [Bacillus halodurans]
 Identities = 194/331 (58%), Positives = 263/331 (78%), Gaps = 1/331 (0%)

Query: 6 ILAVESSCDETSVAILKNDKELLANIIASQVESHKRFGGVVPEVASRHHVEVVTTCFEDA 65
 ILA+E+SCDETS A+++N +L+N+++SQ++SHKRFGGVVPE+ASRHHVE +T E+A

-1780-

Sbjct: 12 ILAIETSCDETSAAVIENGTTILSNVSSQIDSHKRFGGVPEIASRHHVEQITVIVEEA 71

Query: 66 LQEAGIVASDLDAVAVTYGPGLVGALLVGMMAAKAFAWANKLPLIPINHMAGHLMAARDV 125
+ EAG+ +DL AVAVT GPGLVGALL+G+ AAKA A+A++LELI ++H+AGH+ A R +

5 Sbjct: 72 MHEAGVDFADLAAVAVTEGPGLVGALLIGVNAAKAIAFAHQLEPLIGVHHIAGHIYANRL 131

Query: 126 KELQYPLLLLVSGGHTLVVYVSEPGDYKIVGETRDDAVGEAYDKVGRVMGLTYPAGREI 185
KEL++PLLL+VSGGHTLVY+ G++++GETRDDAVGEAYDKV R +GL YP G I

10 Sbjct: 132 KELEFPLLLVVSGGHTELIYMENHGEFEVIGETRDDAVGEAYDKVARTLGLPYPGGPHI 191

Query: 186 DQLAHKGQDITYHFPRAMIKEDHLEFSFSGLSAFINLHHNAEQKGEALVLEDLCASFQAA 245
D+LA G+DT FPRA ++ D +FSFSGLSA IN HNA+Q+GE + ED+ ASFQA+

15 Sbjct: 192 DRLAVNGEDTLQFPRAWLEPDSFDFSFGLSKSAVINTLHNAKQGENVQAEDVAASFQAS 251

Query: 246 VLDILLAKTQKALLKYPVKTLVAGGVAANQGLRERLATDISPD-IDVVIPPLRLCGDNA 304
V+D+L+ KT+KA +Y V+ +++AGGVAAN+GLR L + ID+VIPPL LC DNA

20 Sbjct: 252 VIDVLVTKTKKAAEEYKVRQVLLAGGVAANKGLRTALEEAFKPEIDLVIPLSLCTDNA 311

Query: 305 GMIALAAAEFEKENFASLKLNAKPSLAFES 335
MI AA+I+F+++ FA + LN +PSL E+

Sbjct: 312 AMIGAAASIKFKQOTFAGMDLNGQPSLELEN 342

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4927> which encodes the amino acid sequence <SEQ ID 4928>. Analysis of this protein sequence reveals the following:

25 Possible site: 38
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.76 Transmembrane 86 - 102 (85 - 104)

30 ----- Final Results -----
bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 >GP:BAB04267 GB:AP001508 glycoprotein endopeptidase [Bacillus halodurans]
Identities = 196/330 (59%), Positives = 255/330 (76%), Gaps = 2/330 (0%)

Query: 6 ILAVESSCDETSVAILKNESTLLSNVIASQVESHKRFGGVPEVASRHHVEVITTCFEDA 65
ILA+E+SCDETS A+++N +T+LSNV++SQ++SHKRFGGVPE+ASRHHVE IT E+A

40 Sbjct: 12 ILAIETSCDETSAAVIENGTTILSNVSSQIDSHKRFGGVPEIASRHHVEQITVIVEEA 71

Query: 66 LQEAGISASDLDAVAVTYGPGLVGALLVGLAAAKAFAWANHLPLIPVNHMAGHLMAAREQ 125
+ EAG+ +DL+AVAVT GPGLVGALL+G+ AAKA A+A+ LPLI V+H+AGH+ A R

45 Sbjct: 72 MHEAGVDFADLAAVAVTEGPGLVGALLIGVNAAKAIAFAHQLEPLIGVHHIAGHIYANRL 131

Query: 126 KPLVYPLIALLVSGGHTLVVYVPEPGDYHIIGETRDDAVGEAYDKVGRVMGLTYPAGREI 185
K L +PL+AL+VSGGHTLVY+ G++ +IGETRDDAVGEAYDKV R +GL YP G I

50 Sbjct: 132 KELEFPLLLVVSGGHTELIYMENHGEFEVIGETRDDAVGEAYDKVARTLGLPYPGGPHI 191

Query: 186 DQLAHKGQDITYHFPRAMITEDHLEFSFSGLSAFINLHHNAKQKGDLEILEDLCASFQAA 245
D+LA G+DT FPRA + D +FSFSGLSA IN HNAKQ+G+ + ED+ ASFQA+

55 Sbjct: 192 DRLAVNGEDTLQFPRAWLEPDSFDFSFGLSKSAVINTLHNAKQGENVQAEDVAASFQAS 251

Query: 246 VLDILLAKTKKALSRYPAKMLVAGGVAANQGLRDLAQEI--THIEVVIPKLRLCGDNA 303
V+D+L+ KTKKA Y + +++AGGVAAN+GLR L + I++VIP L LC DNA

60 Sbjct: 252 VIDVLVTKTKKAAEEYKVRQVLLAGGVAANKGLRTALEEAFKPEIDLVIPLSLCTDNA 311

Query: 304 GMIALAAAEYDKQHFANMSLNAKPSLAFD 333
MI AA+I++ +Q FA M LN +PSL +

Sbjct: 312 AMIGAAASIKFKQOTFAGMDLNGQPSLELE 341

An alignment of the GAS and GBS proteins is shown below.

Identities = 288/334 (86%), Positives = 313/334 (93%), Gaps = 1/334 (0%)

-1781-

Query: 1 MKDRYLAVESSCDETSVAILKNDKELLANIIASQVESHKRFGGVVPEVASRHHVEVVT 60
 M DRYLAVESSCDETSVAILKN+ LL+N+IASQVESHKRFGGVVPEVASRHHVEV+TT
 Sbjct: 1 MTDRLAVESSCDETSVAILKNESTLLSNVIASQVESHKRFGGVVPEVASRHHVEVITT 60

5 Query: 61 CFEDALQEAGIVASDLDAVAVTYGPGLVGALLVGMMAAKAFWANKLEPLIPINHMAGHLM 120
 CFEDALQEAGI ASDL AVAVTYGPGLVGALLVG+AAAKAFWAN LPLIP+NHMAGHLM
 Sbjct: 61 CFEDALQEAGISASDLSAVAVTYGPGLVGALLVGLAAAKAFWANHLPLIPVNHMAGHLM 120

10 Query: 121 AARDVKELQYPLIALLVSGGHTLVVYVSEPGDYKIVGETRDDAVGEAYDKVGRVMGLTYP 180
 AAR+ K L YPL+ALLVSGGHTLVVYV EPGDY I+GETRDDAVGEAYDKVGRVMGLTYP
 Sbjct: 121 AAREQKPLVYPLIALLVSGGHTLVVYVPEPGDYHIIGETRDDAVGEAYDKVGRVMGLTYP 180

15 Query: 181 AGREIDQLAHKGQDTYHFPRAMIKEDHLEFSFSGLSAFINLHHNAEQKGEALVLEDLCA 240
 AGREIDQLAHKGQDTYHFPRAMI EDHLEFSFSGLSAFINLHHNA+QKG+ L+LEDLCA
 Sbjct: 181 AGREIDQLAHKGQDTYHFPRAMITEDHLEFSFSGLSAFINLHHNAKQGELELLEDLCA 240

20 Query: 241 SFQAAVLDILLAKTQKALKYPVKTLLVAGGVAANQGLRERLATDISPDIDVVIPLRLC 300
 SFQAAVLDILLAKT+KAL +YP K LVVAGGVAANQGLR+RLA +I+ I+VVIP LRLC
 Sbjct: 241 SFQAAVLDILLAKTKALSRYPAKMLVVAGGVAANQGLRDLAQEIT-HIEVVIPKLRLC 299

Query: 301 GDNAGMIALAAAIIEFEKENFASLKINAKPSLAFE 334
 GDNAGMIALAAAIIE++K++FA++ INAKPSLAF+
 Sbjct: 300 GDNAGMIALAAAIIEYDKQHFANMSLNAKPSLAFD 333

25 SEQ ID 4926 (GBS371) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 7; MW 41kDa), in Figure 170 (lane 4 & 5; MW 55kDa) and in Figure 239 (lane 6; MW 55kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 7; MW 65kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1594

A DNA sequence (GBSx1688) was identified in *S.agalactiae* <SEQ ID 4929> which encodes the amino acid sequence <SEQ ID 4930>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1027(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1595

A DNA sequence (GBSx1689) was identified in *S.agalactiae* <SEQ ID 4931> which encodes the amino acid sequence <SEQ ID 4932>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1307(Affirmative) < succ>

-1782-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1596

- 10 A DNA sequence (GBSx1690) was identified in *S.agalactiae* <SEQ ID 4933> which encodes the amino acid sequence <SEQ ID 4934>. This protein is predicted to be L4171-60 protein. Analysis of this protein sequence reveals the following:

Possible site: 36
>>> Seems to have a cleavable N-term signal seq.

- 15 ----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 20 A related GBS nucleic acid sequence <SEQ ID 10135> which encodes amino acid sequence <SEQ ID 10136> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- >GP:AAC24656 GB:AE001274 L4171.5 [Leishmania major]
Identities = 118/282 (41%), Positives = 167/282 (58%), Gaps = 4/282 (1%)
25 Query: 2 GGTQTNQVVISMLASYEGVIAAETGHVSSHEAGAIEFSGHKVLTLP SHNGKLLASEVAT 61
GGTQTN + S L +E VIA + GH+S+HE GAIE +GHKV+T P +GKL ++
Sbjct: 74 GGTQTNLIACSLALRPWEAVIATQLGHISTHETGAIEATGHKVVTAPCPD GKL RVAD --- 130
30 Query: 62 YIETFYADGNYQHVMVFPGMVYISHPTGYTLYSKAELEELS KICKHYQIPLFIDGARLG Y 121
IE+ + +H MV P +VYIS+ TE GT Y+K ELE++S CK + + LF+DGARL
Sbjct: 131 -IESALHENRSEH MVIPKL VYISNTTEVGTQYTKQELEDISASCKEHGLYLF LDGARLAS 189
35 Query: 122 GLAAKDTDVDFTTAAALSDVFIYIGTKMGALAGEAVVFTKKNRPKQFTTIVKQH GALLAK 181
L++ D+ IA L+D+FYIG TK G + GEA++ ++KQ GAL+AK
Sbjct: 190 ALSSPVNDLTLADIARLTDMFYIGATKAGGMFGEALII LNDALKPNARHLIKQRGALMAK 249
40 Query: 182 GRLLGLAFDRFFTDNLYLKIGKHAIDLABELKIILEEKGY SFYLSPTNQOFIIVENTKL 241
G LLG+ F+ DNL+ ++G H+ +A LK LE G S +NQ F I+ENT +
Sbjct: 250 GWLLGIQFEVLMKDNLFELGAHSNKMAATILKAGLEACGIRLAWPSASNQLFP ILENTMI 309
45 Query: 242 ADLAKNVAYSFW EKYDDHHTVIRLATSWSTSREDVTALRNVL 283
A+L + E D ++RL TSW+T ++ VL
Sbjct: 310 AELNND FDMYTV EPLKDGTCIMRLCTSWATEEKECHRFVEVL 351

No corresponding DNA sequence was identified in *S.pyogenes*.

- 50 SEQ ID 4934 (GBS648) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 8-10; MW 60kDa) and in Figure 186 (lane 6; MW 60kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 12; MW 35kDa), in Figure 140 (lane 10; MW 35kDa) and in Figure 178 (lane 7; MW 35kDa).

Purified GBS648-GST is shown in Figure 243, lane 6; purified GBS648-His is shown in Fig. 229, lane 7.

-1783-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1597

A DNA sequence (GBSx1691) was identified in *S.agalactiae* <SEQ ID 4935> which encodes the amino acid sequence <SEQ ID 4936>. Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.2279(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1598

A DNA sequence (GBSx1692) was identified in *S.agalactiae* <SEQ ID 4937> which encodes the amino acid sequence <SEQ ID 4938>. This protein is predicted to be ribosomal protein S14 (rpsN). Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3848(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12716 GB:Z99108 similar to ribosomal protein S14 [Bacillus subtilis]
Identities = 67/89 (75%), Positives = 76/89 (85%)

35 Query: 1 MAKKSIAKFQKQKLVEQYAE LRRELKEKGDYEALRKLPKDSNPNRLKNRDLIDGRPHA 60
 MAKKS+AK K+Q+LVEQYA +RRELKEKGDYEAL KLP+DS P RL NR ++ GRP A
Sbjct: 1 MAKKS+KAKELKRQQLVEQYAGIRRELKEKGDYEALSKLPDSDAPGR LHNRCMVTGRPRA 60

40 Query: 61 YMRKFGMSRINFRNLAYKGQIPGIKKASW 89
 YMRKF MSRI FR LA+KGQIPG+KKASW
Sbjct: 61 YMRKFMSRIAFRELAHKGQIPGVKKASW 89

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4939> which encodes the amino acid sequence <SEQ ID 4940>. Analysis of this protein sequence reveals the following:

45 Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
50 bacterial cytoplasm --- Certainty=0.3799(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-1784-

Identities = 73/89 (82%), Positives = 85/89 (95%)

Query: 1 MAKKSIAKFQKQKLVQYAEALRRELKEKGDYEALRKLPKDSNPRLKNRDLIDGRPHA 60
 MAKKSIAK+QKQ +L+EQYA+LRR+LK KGDYE+LRKLP+DSNPRLKNRD IDGRPHA

Sbjct: 1 MAKKSIAKYQKQLIEQYADLRDLKAKGDYELRKLPKDSNPRLKNRDKIDGRPHA 60

Query: 61 YMRKFGMSRINFRNLAYKGQIPGIKKASW 89
 YMRKFG+SRINFR+LA+KGQ+PG+ KASW

Sbjct: 61 YMRKFGVSRINFRDLAHKGQLPGVTKASW 89

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1599

A DNA sequence (GBSx1693) was identified in *S.agalactiae* <SEQ ID 4941> which encodes the amino acid sequence <SEQ ID 4942>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5183(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB95931 GB:AL359989 galactose-1-phosphate uridylyltransferase

[*Streptomyces coelicolor* A3(2)]

Identities = 31/105 (29%), Positives = 51/105 (48%), Gaps = 4/105 (3%)

Query: 27 DKCPFC--DKSQLGKILDVKKDDMIWVENKYPTL--EETYQTLVIESNDHNGDISVYESK 82

D+CP C D +L +I D D++ EN++P+L + +V ++DH+ + SE +

Sbjct: 68 DQCPLCPDGERLSEIPDSAYDVVVFENRFPDLAGDSGRCEVVCFTSDHDASFADLSEEQ 127

Query: 83 MRQLLDYLLSKWQLMEESEGHYRSVVLRYNFGPLSGGSLRHPHSQI 127

R +LD + + V + N G G +L HPH QI

Sbjct: 128 ARLVLDAWTDRTSELSHLPSVEQVFCFENRGAEIGVTLGHPHGQI 172

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1600

A DNA sequence (GBSx1694) was identified in *S.agalactiae* <SEQ ID 4943> which encodes the amino acid sequence <SEQ ID 4944>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10133> which encodes amino acid sequence <SEQ ID 10134> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1785-

>GP:EAB06998 GB:AP001518 unknown conserved protein [Bacillus halodurans]
Identities = 186/410 (45%), Positives = 258/410 (62%), Gaps = 27/410 (6%)

5 Query: 4 YDTIIIGGGPAGMMAAISSNFYGNKTLLEIEKNKRLGKLAGTGGGRCNVTNNGNLDELLA 63
++ I+IGGGPAG+MA++S+ +G + LL++K +LG+KLA +GGGRCNVTN LDEL+A
Sbjct: 2 HEVIVIGGGPAGLMASVSAEHGAVLLLDKGDGLGRKLAISGGGRCNVTNRMPLELLA 61

10 Query: 64 GIPGNGRFLYSVFSQFDNHDIIINFFQDNGVTLKEEDHGRMFPTTDKSRTIINALENKIKE 123
IPGNGRF+YS FS F+N DII FF+ G+ LKEED GRMFP +DK+ T++ L +I +
Sbjct: 62 HIPGNGRFMYSPFVSFNNEDIIRFFERLGIALLKEEDGRMFPVSDKATTVVQTLLKRIND 121

15 Query: 124 LGGQIMTDTEVVSVK-KIGDSFYIKTKDTQFASDK-LIVTTGGKSYPTSTGSGFGHDIA 181
LG + T+T V S++ G ++ K+ + K +IV TGG+S P TGGTG + A+
Sbjct: 122 LGVTVRTNTAVASLEYDDGRIAMVQLKNGERLKTKTIVIVATGGQSVFHTGSGTGDAYPWAK 181

20 Query: 182 HFKLEVTDMEEAESPLLTDFP---HKKLQGISLDDVTLSEF---EKHIITH--DLLFTHF 232
+T++ E P+ + P KKLQG+SL D+ LS K I TH D++FTHF
Sbjct: 182 AAGHTITELYPTVPITSAEPFIQEKKLQGLSLRDIELSVYAPNGKQIKTHDGDIMFTHF 241

25 Query: 233 GLSGPAALRISSFVKGETIY-----LDVLPNISVKEL-EIHFQN---EREKSLKNA 280
GLSGPAALR S +V Y +D+ P I + L + QN E +K+LK
Sbjct: 242 GLSGPAALRCSQYVVKALKKYKQPTIEMRIDLRPTIPAEALQETIQNIKAEPKALKITV 301

30 Query: 281 LKILLPERLAEFYAEDL--PEKVQVSVKD--LEMLIQKLKKLPILVTGKMSLAKSFVTK 336
L+ + PER ++ E L + SV+ + ++Q+LK V G +S+ K+FVT
Sbjct: 302 LRGIAPERFLQYIYERLRIDSNLPCASVRHEVIREIVQQLKSFSFHVNGTSLIEKAFVTG 361

Query: 337 GGVDLKEINPKTLESKKVAGLHFAEVLVDINAHTGGFNITSALCTGWVAG 386
GGV +KEI PKT+ SKK AGL F GEVLDI+ +TGG+NIT A TG+ AG
Sbjct: 362 GGVSVKEIEPKTMHKKKAGLFFCGEVLDIHGYTGGYNITCAFSSTGYTAG 411

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4945> which encodes the amino acid sequence <SEQ ID 4946>. Analysis of this protein sequence reveals the following:

35 Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0448(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 308/386 (79%), Positives = 344/386 (88%)

45 Query: 1 MKHYDTIIIGGGPAGMMAAISSNFYGNKTLLEIEKNKRLGKLAGTGGGRCNVTNNGNLDE 60
M YDTIIIGGGPAGMMAAISS++YG KTLLEIEKN+RLGKLAGTGGGRCNVTN+GNLD
Sbjct: 1 MTQYDTIIIGGGPAGMMAAISSSYGYKTLLEIEKNRRLGKLAGTGGGRCNVTNSGNLVD 60

50 Query: 61 LLAGIPGNGRFLYSVFSQFDNHDIIINFFQDNGVTLKEEDHGRMFPTTDKSRTIINALENK 120
L+AGIPGNGRFLYSVFSQFDNHDII FF++NGV LKEEDHGRMFPTTDKSRTII+ALE K
Sbjct: 61 LMAGIPGNGRFLYSVFSQFDNHDIIAFFEENGVLKKEEDHGRMFPTTDKSRTIIDALEKK 120

55 Query: 121 IKELGGQIMTDTEVVSVKKIGDSFYIKTKDTQFASDKLIVTTGGKSYPTSTGSGFGHDIA 180
IK LGGQ++T TEVVSVKK D FY+K+ D F KLIVTTGGKSYPTSTGSGFGHDIA
Sbjct: 121 IKALGGQVLTSTEVVSVKQDDLFLYLSADQFTTCQKLIVTTGGKSYPTSTGSGFGHDIA 180

60 Query: 181 RHFLEVTDMEEAESPLLTDFPHKKLQGISLDDVTLSEFKEHIITHDLLFTHFGLSGPAAL 240
RHFKL VTD+EAESPLLTDFPHK LQGISLDDVTL++KH+ITHDLLFTHFGLSGPAAL
Sbjct: 181 RHFKLTVTDLEAESPLLTDFPHKVLQGISLDDVTLSDKHVITHDLLFTHFGLSGPAAL 240

65 Query: 241 RISSFVKGETIYLDVLPNISVKELEIHFQNEREKSLKNAKILLPERLAEFYAEDLPEK 300
R+SSFVKGGE LD LP++S +L + ++R+K++KNALK LLPER+A+F +ED PEK
Sbjct: 241 RLSSFVKGETIABLDLPLHSTDDLTAYLSDQDKNIKNAKGLLPERVAFLSEDPYEPK 300

Query: 301 VKQVSVKDLEMLIQKLKKLPILVTGKMSLAKSFVTKGGVDLKEINPKTLESKKVAGLHFA 360

-1786-

VKQ+S K + L+ KLK L I +TGKMSLAKSFVTKGGVDLKEINPKTLESKKV GL+FA
 Sbjct: 301 VKQLSPKQEKELLDKLKHLQIPITGKMSLAKSFVTKGGVDLKEINPKTLESKKVPGLYFA 360

Query: 361 GEVLDINAHTGGFNITSALCTGWVAG 386
 GEVLDINAHTGGFNITSALC+GW+AG

5 Sbjct: 361 GEVLDINAHTGGFNITSALCSGWIAG 386

SEQ ID 4944 (GBS196) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 3; MW 44.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 4; MW 69.5kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1601

A DNA sequence (GBSx1695) was identified in *S.agalactiae* <SEQ ID 4947> which encodes the amino acid sequence <SEQ ID 4948>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1550(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10131> which encodes amino acid sequence <SEQ ID 10132> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA73267 GB:Y12736 orfX [Lactococcus lactis subsp. cremoris]
 Identities = 51/173 (29%), Positives = 87/173 (49%), Gaps = 20/173 (11%)

Query: 19 KTVSELAIEILGVSRQAMNNRV-KTLPEECVEK---NSKGVTVVNRDGLIKLBEIYKKTIL 74
 KT+ ELA+ LGVS+Q + N++ K E+ V+ V+N G + KKT+

Sbjct: 6 KTIKELADELGVSKQTIRNKIDKDFREKFVQTIKIKGNNTLVINNAGY----SLLKKTILQ 61

Query: 75 EEEPIDEEASRRELEILVDEKNTEITRLYEQLKAKDIQASKDEQLHVKDIQIAEKDKQ 134
 + + + + + I L EQL K+ Q++ KD+QL KD QI++

Sbjct: 62 NDTAQTAKTLQNDTAQTKL-----ICFLEQLDKKEQQLSVKDKQLENKDTQISQMQL 115

Query: 135 LDQQQQLTLTAMEDTQRLQLELNEAKA-----EVEEIQEAKKEKIQELEAVK 181
 LDQQQ+L L + + + E+NE KA +++++ + E +E+E +K

40 Sbjct: 116 LDQQQLALQDKKLLEEYKSEINELKALKMPREDMKDGSSIRGEAQEETIERLK 168

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4949> which encodes the amino acid sequence <SEQ ID 4950>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3951(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 132/194 (68%), Positives = 154/194 (79%), Gaps = 4/194 (2%)

-1787-

Query: 1 MIFFYKKI---STKEEVMTEKTVSELAELGVSRQAMNNRVKTLPEECVEKNSKGVTVV 57
 M+ F +I S KEE M +EKTVSELA+ILGVSRQA+NNRVK+LPEE ++KN KGVTVV
 Sbjct: 1 MVLFLIRIFSDSDKEENMGIEKTVSELAELGVSRQAVNNRVKSLPEEDLDKNEKGVTVV 60

5 Query: 58 NRDLIKLEEIYKKTILEEEPIDEEASRRELLEILVDEKNTETRLYEQLKAKDIQIASK 117
 R GL+KLEEIYKKTII ++EPI EE +RELLEILVDEKNTETRLYEQLKAKD Q+ASK
 Sbjct: 61 KRSGLVKLEEIYKKTIFDDEPISEETKQRELLEILVDEKNTETRLYEQLKAKDAQIASK 120

10 Query: 118 DEQLHVKDIQIAEKDKQLDQQQQLTAMEDTQRLQLELNEAKAEVEEIQEAKBEKIQEL 177
 DEQ+ VKD+QIAEKDKQLDQQQQLT AM D + L+LEL EAKAE + + + E++Q
 Sbjct: 121 DEQMRVKDVQIAEKDKQLDQQQQLTAKAMADKETLKLELEAKAEANQAR-LQVEEVQAE 179

Query: 178 EAVKKSFFGRFFNK 191
 KK FF R F K

15 Sbjct: 180 VGEKKGFFTRLFAK 193

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1602

20 A DNA sequence (GBSx1697) was identified in *S.agalactiae* <SEQ ID 4951> which encodes the amino acid sequence <SEQ ID 4952>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2157(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06137 GB:AP001515 DNA polymerase III (alpha subunit)
 [Bacillus halodurans]
 Identities = 31/87 (35%), Positives = 52/87 (59%), Gaps = 1/87 (1%)

35 Query: 13 EYIAPDLEFNTVGE-HSHIIQVSAVKYSNHQEIALFDTYVHTKVPLQSFINGLTGITARD 71
 E++ FD+E + ++ II+++AVK N + I F+ + PL + I LTGIT
 Sbjct: 418 EFVVFVDETGLSAVYNKIIELAAVKVNGEIIDRFERFADPHEPLTNTIIELTGITDDM 477

40 Query: 72 IIGAPKIEIVLTDFQSFVGDTPILIGYN 98
 + G P++E VL +F +F+GD L+ +N
 Sbjct: 478 LKGQPEVEQVLNEFHAFIGDAVLVAHN 504

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4953> which encodes the amino acid sequence <SEQ ID 4954>. Analysis of this protein sequence reveals the following:

45 Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.3427(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 136/200 (68%), Positives = 159/200 (79%)

55 Query: 3 FLGEIMKQLQEYIAPDLEFNTVGEHSHIIQVSAVKYSNHQEIALFDTYVHTKVPLQSFIN 62
 FL E MK L YIAPDLEFNTV + SHIIQVSAVKY +H+E+ FDTYV+T VPLQSFIN
 Sbjct: 9 FLEENMKHLDTYIAPDLEFNTVNDVSHIIQVSAVKYDHHEVDSFDTYVYTDVPLQSFIN 68

-1788-

Query: 63 GLTGITARDIIGAPKIEIVLTDFOFVGDTPFLIGYNGYKSDLPPLLVENGLDLTSQYQVDL 122
 GLTGIT+ I PK+E V+ F++FVG+ FLIGYN KSDLP+L ENGLDL QYQ+DL
 Sbjct: 69 GLTGITSKDIAAEKPKVEEVMAPKFNFGELFLIGYNAQKSDLPILAENGLDLRDQYQIDL 128

5 Query: 123 YDEAFVRRSTDNLGIVNLKLTIVADFLGIGKKAHNSLEDARMTARVYEKFLDLLENKIYL 182
 +DEA+ RRS DLNGI NL+L TVA FLGIGK+ HNSLEDARMTA +Y+ FL+ D NK YL
 Sbjct: 129 FDEAYDRRSADNLGIANLRLQTVATFLGIGKGRHNSLEDARMTAVIYKSFLDTINKAYL 188

10 Query: 183 KQKQEVAVDSPFATLGNLFD 202
 QQ+EV D+PFA LG+ FD
 Sbjct: 189 SQQEEVTTDNPFALGDFFD 208

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 1603

A DNA sequence (GBSx1698) was identified in *Sagalactiae* <SEQ ID 4955> which encodes the amino acid sequence <SEQ ID 4956>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

20 INTEGRAL Likelihood = -12.10 Transmembrane 143 - 159 (136 - 166)
 INTEGRAL Likelihood = -4.73 Transmembrane 169 - 185 (168 - 188)

----- Final Results -----

25 bacterial membrane --- Certainty=0.5840(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB42766 GB:AL049841 transcriptional regulator [Streptomyces
 30 coelicolor A3(2)]
 Identities = 46/141 (32%), Positives = 71/141 (49%), Gaps = 11/141 (7%)

Query: 5 YSTGDLAKEAGVTVRTVQYYDKRGILSPSELSEGRRVYSIADLEKLRQIIYLRDLDFSI 64
 YS G +A AGVTVRT+ +YD G+L PSE S G R YS ADL++L+QI++ R+L F +
 35 Sbjct: 3 YSVGQVAGFAGVTVRTLHHYDDIGLLVPSERSHAGHRRYSADLDRLQQLFYRELGFPL 62

Query: 65 DNIGNLFTEDNASQILELFLQVQIRELRL-----AIDSKDKDKLDKAVNLLKTVEKQD 116
 D + L + A L Q ++ R+ A++ + +NL ++
 40 Sbjct: 63 DEVAALLDDPADPRAHLRRQHELLSARIGKLOKMAAAVEQAMEARSMGINL---TPEEK 119

Query: 117 SKTLGYLSDIVLMEENKRKKG 137
 + G EE + +WG
 Sbjct: 120 FEVFGDFDPDQYEEVRRERWG 140

45 There is also homology to SEQ ID 1712.

SEQ ID 4956 (GBS372) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 8; MW 55kDa).

GBS372-GST was purified as shown in Figure 215, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 50 vaccines or diagnostics.

-1789-

Example 1604

A DNA sequence (GBSx1699) was identified in *S.agalactiae* <SEQ ID 4957> which encodes the amino acid sequence <SEQ ID 4958>. This protein is predicted to be cyclopropane-fatty-acyl-phospholipid synthase (mma2). Analysis of this protein sequence reveals the following:

```

5   Possible site: 53
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3145(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAD07482 GB:AE000557 cyclopropane fatty acid synthase (cfa)
      [Helicobacter pylori 26695]
      Identities = 167/397 (42%), Positives = 254/397 (63%), Gaps = 14/397 (3%)

Query: 2   VMSLLIKQLIKSTFDIPLQVTPNGNIETYNGSNPHVKLKLKNFVSSELSKDPSIVLG 61
      ++ ++K + K + QV + + ++ +P LK+++ S++ KD S+ +
20   Sbjct: 1   MISKFLKSMFKQWKNGDYQVVFWDNSVYRNGEHSFKTLKTIHRPLKFSIDIKDMSLTIA 60

Query: 62   EAVMDGDIETIYGSIQELILSAY-RCGDSFLRNSKFSKLIPKQFHDKKHKSQDIQKHYDIG 120
      EA MDG I+I GS+ E++ S Y + L +K I K + S+I KHYD+G
25   Sbjct: 61   EAYMDGVIDIEGSMDEVMHSLYLQTNRYHLHKHDNAKAIQKPIKES----SNISKHYDLG 116

Query: 121  NDFYKLWLDGMTYSCAYFKHENDSLEQAQLNKVHHILNKLNAQPGKLLDIGCGWGTLI 180
      NDFY +WLD+T++YSCAYFK ++D+L AQL K+ H L KL+ +PG KLLDIGCGWG L
30   Sbjct: 117 NDFYSIWLDETLSYSCAYFKDDDTLHAAQLQKLDHTLKKLHLKPGKLLDIGCGWGYS 176

Query: 181  ITAAKEYGLNATGITLSEEQASFITKRIKEEGLENKVTVLIDYRDI---RETYDYITSV 237
      + AA+EYG GIT+S EQ KR++E GLE+KVT+ + +Y+D+ +D + SV
35   Sbjct: 177 VKAAQBYGAEVMGITISSEQYKQANKRVQELGLEDKVTIKLLNYQDLQDGRLYRFDKVVSV 236

Query: 238  GMFEHVGKENLSQYFQTISKRLNINGLALINGITGQVGNHGSCTNSWINKYIFPGGYIP 297
      GMFEHVGK+NL YF+ + + L G+ L+H I G TN+W++KYIFPGGY+P
40   Sbjct: 237 GMFEHVGKDNLPFYFKVKVKEVLKRGGMFLLHSILCCFEKG----TNAWVDKYIFPGGYLP 292

Query: 298  RLITENLNHIASAGLQIADLEPLRRHYQKTLELWTFKNFHNALPEVQK-THDKRFINMWDLY 356
      L E ++ ++ + E LR HY KTL++W NF++ L +V++ ++D+RFI MWDLY
45   Sbjct: 293 SLREVMSVMSECDPHLLMAESLRIHYAKTLDIWRNFNHNLQVKRLSYDERFIRMWDLY 352

Query: 357  LQSCAASFESGNIDIFQYLLSKGVSKDTMPMTRDYMY 393
      L++CA++F G+ D+FQ LL+ V +T P+T++Y+Y
      Sbjct: 353 LRTCASAFRVGSADLFQQLLTNSVD-NTFPLTKKEYIY 388

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1605

50 A DNA sequence (GBSx1700) was identified in *S.agalactiae* <SEQ ID 4959> which encodes the amino acid sequence <SEQ ID 4960>. Analysis of this protein sequence reveals the following:

```

55   Possible site: 35
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4903(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-1790-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11796 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
Identities = 44/97 (45%), Positives = 60/97 (61%)

Query: 1 MMNMQNMNRQAQKLQKQMEQKQADLAASQFTGKSAQELVTVTFTGDKKLISIDYKEAVVD 60
M NMQ MM+Q QK+QK M + Q +LA G + +VTV G K+++ + KE VVD
Sbjct: 5 MGNMQKMMKQMQKMQDMAKAEKVVVEGTAGGGMVTVKANGQKEILDVVIKEEVVD 64

Query: 61 PEDIETLQDMTTQAINDALSQVDDATKKIMGAFAGKM 97
PEDI+ LQD+ A N+AL +VD+ T + MG F M
Sbjct: 65 PEDIDMLQDLVLAATNEALKKVDEITNETMGQFTKGM 101

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4961> which encodes the amino acid sequence <SEQ ID 4962>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4451(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 84/99 (84%), Positives = 94/99 (94%)

Query: 1 MMNMQNMNRQAQKLQKQMEQKQADLAASQFTGKSAQELVTVTFTGDKKLISIDYKEAVVD 60
MMNMQNM+QAQKLQKQMEQKQADLAA QFTGKSAQ+LVT TFTGDKKL+ ID+KEAVVD
Sbjct: 1 MMNMQNMKQAQKLQKQMEQKQADLAAMQFTGKSAQDLVTATFTGDKKLVGIDFKEAVVD 60

Query: 61 PEDIETLQDMTTQAINDALSQVDDATKKIMGAFAGKMPF 99
PED+ETLQDMTTQAINDAL+Q+D+ TTK +GAFAGK+PF
Sbjct: 61 PEDVETLQDMTTQAINDALTQIDETTKKTLGAFAGKLPF 99

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1606

A DNA sequence (GBSx1701) was identified in *S.agalactiae* <SEQ ID 4963> which encodes the amino acid sequence <SEQ ID 4964>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3963(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1791-

Example 1607

A DNA sequence (GBSx1702) was identified in *S.agalactiae* <SEQ ID 4965> which encodes the amino acid sequence <SEQ ID 4966>. Analysis of this protein sequence reveals the following:

Possible site: 48
 5 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.76 Transmembrane 21 - 37 (19 - 39)
 ----- Final Results -----
 10 bacterial membrane --- Certainty=0.2105 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10129> which encodes amino acid sequence <SEQ ID 10130> was also identified.

15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1608

20 A DNA sequence (GBSx1703) was identified in *S.agalactiae* <SEQ ID 4967> which encodes the amino acid sequence <SEQ ID 4968>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1783 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1609

35 A DNA sequence (GBSx1704) was identified in *S.agalactiae* <SEQ ID 4969> which encodes the amino acid sequence <SEQ ID 4970>. This protein is predicted to be probable 1,4-dihydroxy-2-naphthoate octaprenyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence
 40 INTEGRAL Likelihood = -8.76 Transmembrane 239 - 255 (219 - 260)
 INTEGRAL Likelihood = -8.33 Transmembrane 221 - 237 (219 - 238)
 INTEGRAL Likelihood = -6.74 Transmembrane 91 - 107 (89 - 113)
 INTEGRAL Likelihood = -6.32 Transmembrane 39 - 55 (35 - 59)
 45 INTEGRAL Likelihood = -3.77 Transmembrane 111 - 127 (111 - 132)
 INTEGRAL Likelihood = -2.97 Transmembrane 144 - 160 (143 - 161)
 INTEGRAL Likelihood = -1.28 Transmembrane 275 - 291 (275 - 291)
 INTEGRAL Likelihood = -0.59 Transmembrane 177 - 193 (177 - 193)

-1792-

----- Final Results -----

bacterial membrane --- Certainty=0.4503(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15875 GB:Z99123 alternate gene name: ipa-6d-similar to
 quinone biosynthesis [Bacillus subtilis]

Identities = 75/290 (25%), Positives = 139/290 (47%), Gaps = 15/290 (5%)

Query: 5 IFLELVEMKAKTASVLPFLIGLCFSAYYYNSVHPVYVGLFFVAMFLFNMFVDIWNNDY 64
 I +L TAS +P L+G + +Y +++ + F +++ + +++N Y D+
 Sbjct: 21 ILWQLTRPHTLTASFVPVLLGTVLAMFYVKVDLLFLAMLFSCSLWI-QIATNLFNEYDYF 79

Query: 65 RNAVDL-DYKND'TNIIGRENLSLRQIEVIMASLVITSSMIGLVLSQVGLPPLLWMGLFCF 123
 + +D + I R + + I + + + ++G+ + + L +GL
 Sbjct: 80 KRGLDTAESVGIGGAIVRHGMKPKTILQLALASYGIAILLGVYICASSWWLALIGLVGM 139

Query: 124 GIGVLYSFGPRPLSSLPLGEVFSGLTMGFMISLICVYLNTYQNFSDILNLSKIFLISLP 183
 IG LY+ GP P++ P GE+FSG+ MG + LI ++ T D +N+ I LIS+P
 Sbjct: 140 AIGYLYTGGPLPIAYTTPFGELFSGICMGSVFVLISFFIQT-----DKINMQSI-LISIP 192

Query: 184 NTLWIANLMLANNLCDKEDEKNHRYTLVHYTGIRGGLLLFAISNSIALLAIVFEFLFGL 243
 + + + L+NN+ D EED+K R TL G +G + L A S ++A + +V + G
 Sbjct: 193 IAILVGAINLSNNIRDIEEDKKGGRKTLAILMGHKGAVTLAASFAYIWWVGLVITGA 252

Query: 244 APVTVLLSLLLIPFIYKQTKLLWQKQVKRETFVCAVRILALGSATQVLT 293
 A + + L +P + K Q ++ I+A+ S Q T+
 Sbjct: 253 ASPWLFFVFLSVKPKVQAVKGFVNEMPMN-----MIVAMKSTAQTNTF 296

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1610

A DNA sequence (GBSx1705) was identified in *Sagalactiae* <SEQ ID 4971> which encodes the amino acid sequence <SEQ ID 4972>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.22 Transmembrane 155 - 171 (154 - 171)

----- Final Results -----

bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15200 GB:Z99120 similar to NADH dehydrogenase [Bacillus subtilis]
 Identities = 178/403 (44%), Positives = 249/403 (61%), Gaps = 7/403 (1%)

Query: 3 EILVLGAGYAGLKAVRNLQKQSG--DFHITLVDMNDYHYEATELHEVAAGSQPKKITFP 60
 +I++LGAGY GL V L K G D ITLV+ ++YHYE T +HE +AG+ ++ +
 Sbjct: 7 KIVILGAGYGGMLTMTVTRLTKYVGPNDADITLVNKHNYHYETTMHEASAGTLHHDRCRYQ 66

Query: 61 IKDVINTINKVNFQDEVLRVDAENKTVTVKNNGELHYDYVVVALGFVSETFGIKGAMENA 120
 IKDVIN ++VNF+QD V + + K V + N GEL YDY+V+ LG V ETFGIKG E A
 Sbjct: 67 IKDVINQSRVNFVQDTVKAIDEKKVVLAN-GELQDYDLVIGLGAVPETFGIKGLKEYA 125

Query: 121 LQMTNISQAENIHNHIVNTMKLYRETKDE--NLLKLLVCGAGFTGIELAGAMVDERPKYA 178
 + NI+ + + HI Y ++ + L ++V GAGFTGIE G + P+

-1793-

Sbjct: 126 FPIANINTSRLREHIELQFATYNTEAEKRPDRLTIVVGAGFTGIEFLGELAAARVPELC 185

Query: 179 ALAGVKPEQIEIICVEAATRILPMFDDELAQYGVNLIKDLGINLMLGSMIKEIKPGEVVY 238
 V + IICVEAA +LP FD EL Y V+ +++ G+ +G+ ++E P V

5 Sbjct: 186 KEYDVDRSLVRIICVEAAPTIVLPGFDPPELVYAVHYLEENGVEFKIGTAVQECTPEGVRV 245

Query: 239 GTSKEDEELKSITAGTIIWTTGVSGSPVMGESGFDQRRGRVMVNSDLRDEKYDNVYVIGD 298
 G K+DEE + I + T++W GV G P++ E+GF+ RGRV VN DLR P +DNV+++GD

10 Sbjct: 246 G--KKDEEPEQIKSQTVVWAAGVRGHPIVEEAGFENMRGRVKVNPDLRAFGHDNVFILGD 303

Query: 299 VSAFMDTESGRPFPTTAQIATRMGAHVAKNLLHQIKGEATEDFSYSPQGTVASVGNTHGL 358
 S FM+ ++ RP+P TAQIA + G VAKNL IKG E+F +GTVAS+G + +

Sbjct: 304 SSLFMNEDTERPYPPTAQIAMQQGITVAKNLGRLIKGELEEFKPDIKGTVASLGEHNAV 363

15 Query: 359 GVVGKTKIKKYPASVMKKIIMNKSIVDMGGLKELLAKGRFDLY 401
 GVV K+K PAS MKK+I N+SL +GGL L KG+F +

Sbjct: 364 GVVYGRKLKGTIPASFMKKVIDNRSIFMIGGLGLTLKKGKFKFF 406

There is also homology to SEQ ID 4666.

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1611

A DNA sequence (GBSx1706) was identified in *S.agalactiae* <SEQ ID 4973> which encodes the amino acid sequence <SEQ ID 4974>. This protein is predicted to be cytochrome d ubiquinol oxidase, subunit I (cydA-1). Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.64	Transmembrane	19 - 35 (15 - 38)
INTEGRAL	Likelihood = -5.73	Transmembrane	226 - 242 (222 - 244)
INTEGRAL	Likelihood = -4.94	Transmembrane	130 - 146 (126 - 149)
INTEGRAL	Likelihood = -4.83	Transmembrane	429 - 445 (422 - 446)
INTEGRAL	Likelihood = -3.77	Transmembrane	55 - 71 (53 - 74)
INTEGRAL	Likelihood = -3.56	Transmembrane	342 - 358 (340 - 359)
INTEGRAL	Likelihood = -1.06	Transmembrane	89 - 105 (89 - 106)
INTEGRAL	Likelihood = -0.59	Transmembrane	186 - 202 (186 - 202)

----- Final Results -----

bacterial membrane	---	Certainty=0.3654(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15902 GB:Z99123 cytochrome bd ubiquinol oxidase (subunit I)
 [Bacillus subtilis]

45 Identities = 246/470 (52%), Positives = 319/470 (67%), Gaps = 12/470 (2%)

Query: 6 LARFQFAMTTFVHFFHFVFTTIGTCLVVAIMETMYVITKNEEYKCLKFWGNIMLLSFAVG 65
 LAR QFA TT+FH FVP +IG +VA+MET+Y++ KNE Y K+ KFWG++ L++FAVG

50 Sbjct: 6 LARIQFASTTLFHFPLFVPMISGLVFMVALMETLYLVKKNELYLKMAKFWGHLFLINFAVG 65

Query: 66 VVTGIIQEFQFGMNWSDYSRFGDIFGAPLAIEALLAFFMESTFLGLWMFTWDNKKISKK 125
 VVTGI+QEFQFG+NWSYSRFVGD+FGAPLAIEALLAFFMES F+GLW+F WD ++ KK

Sbjct: 66 VVTGIIQEFQFGLNWSYSRFVGDFGAPLAIEALLAFFMESIFIGLWIFGWD--RLPKK 123

55 Query: 126 LHVTFIWLTVFGSLMSAMWILTANSFMQHPVGYEVVNGRAQMTDFLALVKNPQFFYEFTH 185
 +H IWL V FG++MS+ WILTANSFMQ PVG+ + NGRA+M DF AL+ NPQ + EF H

Sbjct: 124 IHALCIWLVSFGTIMSSFWILTANSFMQEPVGFITKNGRAEMNDFGALITNPQLWVEFPH 183

60 Query: 186 VIFGAITMGGTVVAGMSAFRLKSEQLKDTTVELYKKSVRIGLVALLGSISVMGVGDLQ 245
 VIFGA+ G +AG+SAF+LLK ++ V +K+S ++ ++V L + V G +Q

-1794-

Sbjct: 184 VIFGALATGAFFIAGVSAFKLLKKKE-----VPFFKQSFKLAMIVGLCAGLGVLGSGHMQ 238

Query: 246 MKALIHDPMPKFAAMEGDYEDSGDPAAWSVVAWANEAEHKQVFGIKIPYMLSILSYGKPS 305
+ L+ QPMK AA EG +EDSGDPAAW+ A + K IK+PY LS L+Y K S

5 Sbjct: 239 AEHLMESQPMKMAASEGLWEDSGDPAAWTAFATIDTKNEKSSNEIKVPYALSILAYQKFS 298

Query: 306 GSVKGMdTANKELVAKYGKDNYPVMNLLFYGFRTMAAMGTAIMGVSVLGLFLTRKKKPI 365
GSVKGM T E YGK +Y P V F+ FR M G ++ ++ GL+L R+KK

10 Sbjct: 299 GSVKGMKTLQAEYEKIYKGDIYPPVKTTFWSFRIMVGAGVVMILAALGGLWLNRRKK-- 356

Query: 366 LYKHKWMLWIVALTTTFAPFLANTFGWIVTEQGRYPWTIVYGLFKIKDSVSPNVSVASLFVS 425
L KW L I+ PFLAN+ GWI+TE GR PWTV GL SVSPNV+ SL S

Sbjct: 357 LENSQWYLRIMIALISFPFLANGAGWIMTEIGRQPWTVMGLMTAQSVSPNVTAGSLLFS 416

15 Query: 426 NTVYFLLFGGLAVMMISLTIRELKKGPEYEDEHGHHCAYTSIDPFEEGAY 475
+ +++ L +++ L IRE+KKG E+++ HH S DPF + Y

Sbjct: 417 IIAFGVMYMLGALLVFLFIREIKKGAHDN---HHDVPVSTDPFSSQEVY 463

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1612

A DNA sequence (GBSx1707) was identified in *S.agalactiae* <SEQ ID 4975> which encodes the amino acid sequence <SEQ ID 4976>. This protein is predicted to be cytochrome oxidase subunit II (cydB-1).

- 25 Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -14.49	Transmembrane	226 - 242 (220 - 250)
INTEGRAL	Likelihood = -8.12	Transmembrane	254 - 270 (250 - 282)
INTEGRAL	Likelihood = -7.64	Transmembrane	198 - 214 (196 - 218)
INTEGRAL	Likelihood = -6.95	Transmembrane	85 - 101 (76 - 103)
INTEGRAL	Likelihood = -6.74	Transmembrane	6 - 22 (1 - 27)
INTEGRAL	Likelihood = -6.16	Transmembrane	300 - 316 (298 - 322)
INTEGRAL	Likelihood = -5.36	Transmembrane	119 - 135 (117 - 143)
INTEGRAL	Likelihood = -4.04	Transmembrane	159 - 175 (155 - 178)

30

35

----- Final Results -----

bacterial membrane	---	Certainty=0.6795(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

40

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15901 GB:Z99123 cytochrome bd ubiquinol oxidase (subunit II)
[Bacillus subtilis]

45 Identities = 158/331 (47%), Positives = 223/331 (66%), Gaps = 1/331 (0%)

Query: 1 MSALQFFWFFLIGLLFSGFFFLBGFDFGVGMVAVQTLTHNEHEKDQVETIGPVWDGNEVW 60
M++L WF L+ +LF GFFFLEGFDFGVGMA + L HNE E+ ++ TIGP WD NEVW

Sbjct: 1 MASLHDLWFIWVAVLFVGGFFFLBGFDFGVGMATRLGHNELERRVLINTIGPFWDA NEVW 60

50

Query: 61 LLTGGGAMFASFPYWYASLFSGYYLILLTILFGLIIRGVSEFRHKVPAEK-KQFWNWTIL 119
LLTG GA+FA+FP WYA++ SGYY+ + +L L+ RGV+FEFR KV K + W+W +

Sbjct: 61 LLTGAGATFAAFPNWYATMLSGYYIPFVIVLLALMGRGVAFEFGRKVDHLKVVVWDVWV 120

55

Query: 120 TIGSAIVPFFFGIMFISLIQGMPLDASGNLSAQFSDYFNIFSLVGGVAMVLLAYLHGLNY 179
GS I PF G++F +L +GMP+DA N+ A SDY N++S++GGV + LL + HGL +

Sbjct: 121 FFGSLIPPFLVGLFTTLFRGMPIDADMNIHAHVSDYINVYSILGGVTVTLLCFQHGLMF 180

60

Query: 180 IALKTEGPIREARARNYAQLLYWVLYLGLALFAVLLYFKTDFFSNHPIVTTIMVLVIVVLA 239
I L+T G ++ RAR AQ + V+++ + FA L ++TD F+ +T + ++IV+

-1795-

Sbjct: 181 ITLRTIGDLQNRARKMAQKIMGVVFAVLAFALSAAYQTMFTRRGEITIPLAVLIVICF 240

Query: 240 VLAHASTFKGAEMTAFLASGLSLVSVVLLFQGLFPRVMISSISPKYDLLIONASSTPYT 299

5 +LA K + F +G L V ++F LPPRM+SS+ YDL + NASS Y+
Sbjct: 241 MLAADVIRKKKGWTFGMTAGLALTVMIFISLFPVMVSSLHSAYDLTVANASSGDYS 300

Query: 300 LKVMISIVAITLVFPVFLAYTAWAYYIFRKRIIT 330

LKVMIS A+TL+PFV+ W+YY+FRKR++

10 Sbjct: 301 LKVMISIALTLPLFVIGSQIWSYYVFRKRVS 331

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1613

15 A DNA sequence (GBSx1708) was identified in *S.agalactiae* <SEQ ID 4977> which encodes the amino acid sequence <SEQ ID 4978>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

20 ----- Final Results -----

 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1614

30 A DNA sequence (GBSx1709) was identified in *S.agalactiae* <SEQ ID 4979> which encodes the amino acid sequence <SEQ ID 4980>. This protein is predicted to be transport ATP-binding protein cydc (cydD). Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have an uncleavable N-term signal seq

35 INTEGRAL Likelihood = -16.82 Transmembrane 158 - 174 (144 - 182)

 INTEGRAL Likelihood = -6.48 Transmembrane 15 - 31 (14 - 34)

 INTEGRAL Likelihood = -5.31 Transmembrane 243 - 259 (238 - 266)

 INTEGRAL Likelihood = -2.55 Transmembrane 136 - 152 (134 - 152)

40 INTEGRAL Likelihood = -0.48 Transmembrane 263 - 279 (263 - 279)

----- Final Results -----

 bacterial membrane --- Certainty=0.7729(Affirmative) < succ>

 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15900 GB:Z99123 ABC membrane transporter (ATP-binding protein) [Bacillus subtilis]

50 Identities = 279/569 (49%), Positives = 401/569 (70%), Gaps = 6/569 (1%)

Query: 2 LDKAVMRLSGIHKLGLLAGLDVLQAIIFIIGQAYYLSLSITGLWEGQKLSSQTVYILLEFM 61

+ K + R G+ ++L L+ L ++Q II QA +LS ++TGL+ G+ ++S I F+

Sbjct: 1 MGKDLFRYKGMKRILTLITCLTIQTAAIIMQAEWLSEAVTGLFNGKGITSLLEPVIGFFL 60

-1796-

Query: 62 VSYLGRHVIDYIKNRKLDDEFSTAQSSLLRRQLLDKLFDLGPKVVQEQTGNVVTMALDGV 121
 +++++ RH + + + + + LR+ LD+LF LGP+ +++GTG +VT+A++G+
 Sbjct: 61 IAFIARHGMTVARQKIVYQYAARTGADLRKSFLDQLFRLGPRFAKKEGTGMVTLAMEGI 120

Query: 122 SLVENYLRVLNKMINSIIPWIIAYIFYLDIESGAILLIVFPLIIIFMIILGYAAQAK 181
 S YL L L KM++M+I+P ++ Y+F+ D S IL+ P++IIFMI+LG AQ K
 Sbjct: 121 SQFRRYLELFLPKMVSMIAIVPAAVVIYVFFQDRTSAILVAAMPILIIIFMILLGLVAQRK 180

Query: 182 ADKQYESYQVLSNHFDSLRLGIDTLKYFGLSKRYGKSIYQTSESFRKATMSTLKIGILST 241
 AD+Q++SYQ LSNHF+DSLRLG++TL++ GLSK + K+I+ SE +RKATMSTL++ LS+
 Sbjct: 181 ADRQWKSYQRLSNHFVDSLRLGLETLRFLGLSKSHSKNIFYVSERYRKATMSTLRVAFLLSS 240

Query: 242 FALDFFTTLSIAIVAVFLGLRLNEQIYLLPALTILILSPEYFLPVRDFSSDYHATLDGK 301
 FALDFFT LS+A VAVFLGLRL++ I L PALT LIL+PEYFLPVR+ +DYHATL+G+
 Sbjct: 241 FALDFFTMLSVATVAVFLGLRLIDGILLGPALTALILAPEYFLPVRREVGN DYHATLNGQ 300

Query: 302 NAFQAIQKVLNKTGIKGE-QLVIDDWSKESRLDLENIAIAYDQKRVEDVTLRFRGHQKV 360
 A + IQ++L++ G K E L ++ WS + L L +++ R V D+ L F+G +K+
 Sbjct: 301 EAGKTIQEILSQPGFKEETPLQLEAWSQDELKLSGVSVG----RSVSDIHLSPFKGKKKI 356

Query: 361 ALVGVSGSGKSSLINLLSGFLGPDNGSLKVDGREVTNLDQEDWHKQMIYIPQTPYVFEMS 420
 ++G SG+GKS+LI++L GFL PD G ++V+G ++L W K ++YIPQ PY+F+ +
 Sbjct: 357 GIIGASGAGKSTLIDILGGFLEPDGGMIEVNGTSRSHLQDGSWQKNLLYIPQHPYIFDDT 416

Query: 421 LRDNITFYTPNASDEEVVRAIHMVGLDLSLSELPDGLETRIGNGARPLSGGQAQRALAL 480
 L +NI FY P+AS E+ RA GL L++ LPDGLE RIG G R LSGGQAQR+ALAL
 Sbjct: 417 LGNNIRFYHPSASAEDTTTAAASAGLTENVNLPDGLEGRIGEGGRALSGGQAQRVALAL 476

Query: 481 AFLDQNRIRMVDFDEPTAHLDDIETELELKEKMLPLMSDRLVIFATHRLHLWNQMDVIVVME 540
 AFL NR I++ DEPTAHLDDIETE E+KE ML L D+LV ATHRLHW+ MD I+V++
 Sbjct: 477 AFLG-NRPILLDEPTAHLDDIETEYIKETMLDLFEDKLVFLATHRLHWMLDMEIIVLD 535

Query: 541 KGRVAEVSQYQELLAKKGYLYQLKHAMGG 569
 GRVAE+G++ ELL K G +L A G
 Sbjct: 536 GGRVAEIGTHNELLEKNGVYTKLVKAQLG 564

A related DNA sequence was identified in *Spyogenes* <SEQ ID 4981> which encodes the amino acid sequence <SEQ ID 4982>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.61	Transmembrane	159 - 175 (154 - 190)
INTEGRAL	Likelihood = -10.03	Transmembrane	70 - 86 (63 - 91)
INTEGRAL	Likelihood = -3.03	Transmembrane	282 - 298 (282 - 301)
INTEGRAL	Likelihood = -1.44	Transmembrane	261 - 277 (260 - 278)

----- Final Results -----
 bacterial membrane --- Certainty=0.5246(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC22320 GB:U32749 ATP-binding transport protein (cydD)
 [Haemophilus influenzae Rd]
 Identities = 167/544 (30%), Positives = 279/544 (50%), Gaps = 15/544 (2%)

Query: 46 MISFYLIAKTFSTFILGHAIALGRIAGLLLLLNVGFLAILGK---QLQGIASQFARDS 102
 + S+ L A F L A+ LG + L L A GK Q AS +
 Sbjct: 17 VFSYILQAYFHELSSLSAVILGIVLIAALALR-----AFAGKKSQASYFASTKVKHE 70

Query: 103 LKQSFFFEAFIDLDGQFDAHASDADILTASQGIDSLDTYYGYLL-SLSMRKWNCTTIMI 161
 L+ + + S + I+ +AS+G++ I+ Y+G YL L T
 Sbjct: 71 LRSLIYRKLASMPLNQVNQQSTSSIIQVASEGVEQLEIYFGRYLPQLFYSLAPLTLFAF 130

Query: 162 LVFLIYPLAGLVFLGVLPPLIPLSIVAMQKRSQPNMSHYSSYMDVGNLFMDDLKGLNTLY 221

-1797-

L+F + A ++ L +PLIP+SI+A+ K ++ ++ YWS Y+ +G+ F+D+L+GL TL
 Sbjct: 131 LIFFSFKTA-IILLICVPLIPMSIIAVNKIAXKLLAKYWSIYVGLSSFLDNLQGLITLK 189

5 Query: 222 SYQATERYEQEFGKAEQFRKATMSLLGFQLQAVGYMDAVMYLGIGLSGFLAVQALATGQ 281
 YQ + +AE FRK TM +L QL +V MD + Y G + A+ Q
 Sbjct: 190 IYQDDAYKAKAMDKEAEHFRKITMKVLTMLNSVSLMDLLAYGGAIGILTALLQFQNAQ 249

10 Query: 282 LSFFNFFLLIATEFFTPIREQGYGMHLVMMNTKMADRIFSFLDS-VPARKDNKSKITAI 340
 LS + F+L+++EFF P+R G H+ M +D+IF+ LD+ V ++ A
 Sbjct: 250 LSVLGVILFILLSSEFFIPLRLGSHFFHVAMNGKAASDKIFTLLDTPVETQSAVDPEAK 309

15 Query: 341 NFNQIDIQNISLAY-EKKTVLSGVMTTLTKGQLTAIAGVSGQKTSLAQLLLKRQSATTG 399
 N Q++I+++ +Y E+K ++G+ +++ QL+ G SG GK++L LL+ A G
 Sbjct: 310 NNVQVEIKDLHFSYSEKPAITGLNLSILPNQLSVFVGKSGCGKSTLVSLLMGFNKAQQG 369

20 Query: 400 HILFDGLSDNLSQETINQQVLYVSDQSTLLNRSIYDNLRLA-ANLSKKEILDWIDQHGL 458
 ILF+G ++ N+ + + Q+V VS S + ++ +N+ +A + + ++I ++Q L
 Sbjct: 370 EILFNGQNALNIDRTSFYQKVSLSHSSVVFKGTLRENMTMAKIDATDEQIYACLEQVNL 429

25 Query: 519 DNLITRLAKTAIVIVITHKMSRLKGANQVLFNLNGQPACLGKPCDLYRDQPTYRHLVDTQ 578
 I + + +++I+H+++ A+ + L+ G+ G +L Q Y + Q
 Sbjct: 489 LQFIQQKQKKTIVMISHRLANAVNADCINVLDQKGLIEQGTHTKELMEKQGAYAEFQQQ 548

30 Query: 579 ARLE 582
 LE
 Sbjct: 549 KDLE 552

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/552 (25%), Positives = 260/552 (46%), Gaps = 12/552 (2%)

35 Query: 1 MLDKAVMRLSGIHKLLGLLAGLDVLQAFIIGQAYYLSLSITGLWEGQKLSSQTVYILLF 60
 +L + R++ LL + A L LQ + + Y ++ + + G ++ + LL
 Sbjct: 16 LLKRLRERAPKRYLLYVSAPLSWLQFVMRMISFYLIAKTFSTFILGHATAIGRLAGLL 75

40 Query: 61 MVSYLGRHVLDYIKNRKLDLDFSTAQSSLLRRQLDLKFLDLPKVVQEQGTGNVVIMALDG 120
 +++ +G V+ + + S L++ + DL + +++T+A G
 Sbjct: 76 LLNVVG-FVLAILGKQLQGIASQFARDSLKQSFEAFIDLDGQFDAHASDADILTASQG 134

45 Query: 121 VSLVENYLRVLNKMINSIIPWIIILAYIFYLDIESGAILLIVFPLIIIFMIILGYAAQA 180
 + ++ Y L+ + I+ +F + +G + L V PLI + ++ + +Q
 Sbjct: 135 IDSLDTYGYYSLSMRKTWNCTTIMILVFLIYPLAGLVFLGVPLIPLSIVAMQKRSQP 194

50 Query: 181 KADKQYBSYQVLSNHFLDSLRLGIDTLKYFGLSKRYGKSIYQTSSESFRKATMSTLKIGILS 240
 + SY + N F+D L+G++TL + ++RY + +E FRKATMS L + +
 Sbjct: 195 NMSHYWSSYMDVGNLFDLKLNTLYSYQATERYEQEFGKAEQFRKATMSLLGFQLQA 254

55 Query: 241 TFALDFFTTLSIAIVAVFLGLRLNEQIYLLPALTILILSPEYFLPVRDFSSDYHATLDG 300
 +D L I + L Q+ L L+++ E+F P+R+ H +
 Sbjct: 255 VGYMDAVMYLGIGLSGFLAVQALATGQLSFFNFFLLIATEFFTPIREQGYGMHLVMMN 314

60 Query: 301 KNAFQAIQKVLNKTGIRGEQLVIDDWSKE----SRLDLENIAIAYDQKRVVEDVTLRFRG 356
 I L+ + D+ SK +++D++NI++AY++K V+ VT+
 Sbjct: 315 TKMADRIFSFLDSVPARK-----DNKSKTAINFNQIDIQNISLAYEKKTVLSGVMTMTLK 369

65 Query: 357 HQKVALVGVSGSGKSSLINLLSGFLGPDNGSLKVDGREVINLDQEDWHKQMIYIPQTPYV 416
 Q A+ GVSG GK+SL LL G + DG + NL QE ++Q++Y+ +
 Sbjct: 370 GQLTAIAGVSGQKTSLAQLLLKRQSATTGHILFDGLSDNLSQETINQQVLYVSDQSTL 429

Query: 417 FEMSLRDNITFYTPNASDEEVVRAIHMVGLDLSLSELPDGLETRIGNGARPLSGGQAQRI 476
 S+ DN+ N S +E++ I GL S ++ LPDGL+T +G LS GQ Q++
 Sbjct: 430 LNRSIYDNLRL-AANLSKKEILDWIDQHGLLSFINWLPDGLDTIVGENGILLSGKQKQV 488

Query: 477 ALARAFLDQNRIMVFDEPTAHLDIETELKEKMLPLMSDRLVIFATHRLHWNQMDVI 536

-1798-

ARA L + R + +FDE T+ LD E E + + L +VI TH++ L + +
 Sbjct: 489 ICARALLSK-RSLYIFDEATSSSLDAENERIIDNLITRLAKTAIVIVITTHKMSRLKGANQV 547

Query: 537 VVMEKGRVAEVG 548

+ + G+ A +G

Sbjct: 548 LFLNTGQPACLG 559

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 1615

A DNA sequence (GBSx1710) was identified in *Sagalactiae* <SEQ ID 4983> which encodes the amino acid sequence <SEQ ID 4984>. This protein is predicted to be transport ATP-binding protein cydd (cydC). Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.84	Transmembrane	260 - 276 (258 - 284)
INTEGRAL	Likelihood = -9.34	Transmembrane	172 - 188 (147 - 199)
INTEGRAL	Likelihood = -6.53	Transmembrane	150 - 166 (147 - 171)
INTEGRAL	Likelihood = -6.05	Transmembrane	31 - 47 (29 - 52)
INTEGRAL	Likelihood = -3.35	Transmembrane	68 - 84 (67 - 84)
INTEGRAL	Likelihood = -1.17	Transmembrane	293 - 309 (292 - 310)
INTEGRAL	Likelihood = -0.69	Transmembrane	494 - 510 (493 - 510)

----- Final Results -----

bacterial membrane --- Certainty=0.6137(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10127> which encodes amino acid sequence <SEQ ID 10128> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15899 GB:Z99123 ABC membrane transporter (ATP-binding protein) [Bacillus subtilis]
 Identities = 262/573 (45%), Positives = 389/573 (67%), Gaps = 14/573 (2%)

Query: 16 LKTDQWIKPFFKQYKVSIVIALFLGFMFTFFSASALMFNSGYLISKASASLPNILLVYVPI 75
 +K ++WI P+ KQ V+ +FLG +T PSA+ LMF SG+LISK+A+ P NILL+YVPI
 Sbjct: 1 MKKEEWILPYIKQONARLFVLVIFLGAVTIFSAFLMFTSGFLISKAATRPENILLIYVPI 60

Query: 76 VLTRAFGIGRPVFRYIERLTSHNWVLRMTSQLRLKLYHSLESNAIFMKRDFRLGDVGMGLL 135
 V R FGI R V RY+ERL H+ +L++ S +R++LY+ LE A+ ++ FR GD++G+L
 Sbjct: 61 VAVRTFGIARSVSRIVERLVGHIIILKIVSDMRVRLYNMLEPGALMLRSRFRFTGDMGLIL 120

Query: 136 AEDINYLQNLVLRITFPTIIAWILYSFIIITATGFFSLWFALMMLLYLAIMIFLFPWLSIL 195
 +EDI +LQ+ +L+TIFP I A +LY+ +IA GFFS FA+++ LYL +++ LFP+ S+L
 Sbjct: 121 SEDIEHLQDAFLKTIFPAISALLLYAVSVIALGFFSWPFAILLALYLFVLVVLFPVVSLL 180

Query: 196 ANGARQTRKELKNHLYTDLTDNVLGISDWIFSQRGQEXVALHERSESELMVQKKIRSF 255
 A+ + K +N LY+ LTD V+G+SDW+FS R ++ +E+ E + +++K + F
 Sbjct: 181 VTRAKNAKLKSGRNVLSRLTDAVMGVSDWMFSGRRHAFIDAYEKEERDWFELERKKQRF 240

Query: 256 DNRRALIVELVFGFLAILVLIWASNQFIGHRGGEA--NWIAAFVLTVPFLSEAFAGLSAA 313
 R + + L +L++ W + Q GE IAAFVL VFPL+EAF LS A
 Sbjct: 241 TRWRDFAAQCLVAGLILLMLFWTAGQ---QADGELAKTMIAAFVLVVFPLTEAFPLPSDA 297

Query: 314 AQETNKYSDSIHLRN-----ELSETYFETTONQLENPKPYDFSVKNLSFQYKQEKWVLH 367
 E Y DSI R+N E S+T E+ L + + +++F Y + VLH
 Sbjct: 298 LGEVPGYQDSIRRMNNVAPQPEASQT--ESGDQILDLDQVTLAFRDVTFSYDNSSQ-VLH 354